



Modélisation 3D de scènes urbaines à partir d'images satellitaires à très haute résolution

Nesrine Chehata

► To cite this version:

Nesrine Chehata. Modélisation 3D de scènes urbaines à partir d'images satellitaires à très haute résolution. Interface homme-machine [cs.HC]. Université René Descartes - Paris V, 2005. Français. NNT: . tel-00011529

HAL Id: tel-00011529

<https://theses.hal.science/tel-00011529>

Submitted on 2 Feb 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université RENÉ DESCARTES - PARIS 5
Centre universitaire des Saints-Pères
UFR DE MATHÉMATIQUES ET INFORMATIQUE

Thèse présentée pour l'obtention du grade de Docteur
de l'université RENÉ DESCARTES-PARIS 5
Spécialité : **Informatique**

Sujet de thèse :

**Modélisation 3D de scènes urbaines à
partir d'images satellitaires à très haute résolution**

NESRINE CHEHATA

Soutenue le mardi 21 Juin 2005, devant le jury composé de :

Pr. Nicole Vincent (université Paris 5) : Présidente du jury
Pr. Georges Stamon (université Paris 5) : Directeur de thèse
Pr. Henri Maître (ENST) : Rapporteur
Dr. Mathieu Cord (ENSEA) : Rapporteur
Dr. Marc Pierrot-Deseilligny (HDR - IGN) : Examineur
Dr. Franck Jung (ESGT) : Examineur
Dr. Gilbert Pauc (CNES)

Remerciements

Je tiens en premier lieu à remercier ma famille, mes parents et mon frère Amine qui m'ont énormément soutenu durant cette thèse et qui m'ont encouragé tout au long de mon parcours. Je leur dédie ce travail qui n'aurait pas eu lieu sans leur soutien.

Je veux exprimer ici ma sincère gratitude au Pr. Georges Stamon, mon directeur de thèse, qui m'a suivi et soutenu tout au long de mes travaux de recherche et qui a su me guider dans les moments de doute. Je lui en suis très reconnaissante.

Je remercie par ailleurs tous les membres du jury : tout d'abord, le Pr. Nicole Vincent de m'avoir fait l'honneur de présider mon jury de thèse. Je remercie également mes rapporteurs, le Pr. Henri Maître et le Dr. Mathieu Cord pour le temps consacré à la lecture de ce manuscrit de thèse et pour leurs remarques pertinentes. Je remercie, par la suite, M. Marc-Pierrot Deseilligny et tout particulièrement Franck Jung, mon encadrant à l'IGN pour son suivi et sa relecture minutieuse de ce manuscrit. Je remercie enfin, le Dr. Gilbert Pauc du CNES d'avoir suivi cette thèse et d'avoir orienté ce travail vers un contexte applicatif.

Mes remerciements s'adressent ensuite à toute l'équipe du SIP-CRIP5 de l'université Paris 5 pour la bonne ambiance qui régnait au sein de l'équipe et les nombreuses discussions qu'on a eu durant ces trois années de thèse. Je remercie tout particulièrement les permanents Florence Cloppet-Oliva et Nicolas Loménie pour leurs conseils et les enseignements pédagogiques qu'ils m'ont apportés.

Je remercie également tous les membres du laboratoire MATIS de l'IGN, qui m'a accueilli durant cette thèse. De nombreuses discussions enrichissantes m'ont permis de mener à bien ce travail. Je suis tout particulièrement reconnaissante à Franck Taillandier, mon collègue de bureau pendant ces trois années pour son suivi, son soutien, sa disponibilité et sa rigueur scientifique. Je remercie également M. Hervé le-Men et Nicolas Paparoditis pour leurs précieux conseils. Je remercie Didier Boldo, chef du laboratoire MATIS, pour sa disponibilité et son aide, Grégoire Maillet et David Flamanc pour leurs avis d'experts en reconstruction 3D et pour tous les outils mis à disposition à l'IGN, Sylvain Airault pour son expertise dans l'évaluation des MNE. Je suis également reconnaissante à François Boyero, pour sa disponibilité, son aide précieuse, sa bonne humeur sans oublier son franc-parler tant apprécié.

Je n'oublie évidemment pas mes collègues de bureau qui ont permis d'avoir une bonne ambiance de travail et surtout une très bonne entente. Par ordre chronologique, je remercie Olivier De-joinville, Frédéric Bretar, Florent Lafarge et Clément Mallet. Les moments qu'on a partagés ensemble resteront un très bon souvenir.

ET, je remercie évidemment tous mes amis qui m'ont soutenu durant ces trois années de recherche et qui ont suivi de près le déroulement de ma thèse avec ses moments de doute, ses rebondissements et ses moments de réussite. Un grand merci à tous.

Summary

This research work is a part of a global project related to urban scene modeling from high resolution satellite images with focus on building reconstruction. This work has been carried out as a collaboration among the IGN (French National Geographical Institute), the CNES (The French National Space Center) and the University of Paris 5. The input data consist of a panchromatic stereo pair of satellite images, with a submetric resolution of 50-70 cm and a low Base to Height ratio B/H [0.05 - 0.2]. Since a detailed extraction and description of building rooves is complex in a satellital context, we propose to describe the scene by means of a 3D surface which provides either raster or vector information and thus different description levels. A "Hypothesis-and-verify" strategy is used. 3D-primitives are extracted from images. Two complementary approaches are proposed which are based on 3D-segments for the first one and on region matching for 3D-facets estimation. The originality of our approach is a global matching of multiscale segmentations which provides reliable 3D-facets.

Primitives validation is done by means of a 3D-surface modeling process. 3D-surface computation can be formulated as an energy minimization process based graph cuts. 3D hybrid graph is constructed on raw correlation information and extracted primitives. Minimal capcity cut will provid the 3D surface model. The main contribution of our approach is the use of 3D-primitives such as 3D-segments and 3D-facets as well as the introduction of information from external database such asroad database or cadastral maps, to guide the optimization process. We propose an improvement over the graph cut method for stereo computation. The main idea is that to insert depth constraints based know 3D primitives. Then the graph cut problem is reformulated to enable better discontinuity computation and surface regularization. The obtained product is an hybrid DEM (Digital Elevation Model) which provides the highest level of reliable primitives for each scene region.

keywords : satellital images, stereoscopy, 3D-surface modeling, 3D-primitives, segmentation, multi-scales region matching, graph cuts optimization, Photogrammetry, Digital Elevation Model.

Résumé

Ce travail s'insère dans un projet global de modélisation de scènes urbaines à partir d'images satellitaires très haute résolution. Le projet a été proposé dans le cadre d'une collaboration entre l'IGN et le CNES et l'université de Paris 5. L'objectif final du projet sera de fournir un système automatique permettant de classifier la scène en deux thèmes ; bâti et routes et de reconstruire les bâtiments en 3D. Dans le cadre de cette thèse, on s'intéresse uniquement à la modélisation des bâtiments. On dispose en entrée de couples stéréoscopiques panchromatiques à [50-70 cm] de résolution et de faible rapport Base sur Hauteur B/H [0.05-0.2]. On exclut la multiscopie. Étant donnée la complexité d'une extraction fine et détaillée des toits en contexte satellitaire, on propose de modéliser la scène urbaine par une surface 3D hybride fournissant différents niveaux de description en fonction de la fiabilité des primitives extraites : points 3D, segments 3D et surfaces planes. Une stratégie mixte a été adoptée. Tout d'abord, une stratégie ascendante basée sur les images permet d'extraire les primitives 3D (segments 3D et facettes 3D). Deux stratégies complémentaires d'extractions de primitives seront détaillées. Une description multi-échelles est utilisée pour la segmentation des images. Notre apport consiste essentiellement dans l'appariement global de deux segmentations multi-échelles du couple stéréoscopique.

L'ensemble des primitives sera validé par une approche descendante et permettra de contraindre la modélisation de la surface 3D. Le problème de modélisation de surface 3D peut être formulé comme un problème de minimisation d'énergie. Il sera résolu par optimisation à base de flots de graphes, contrainte par les primitives 3D. Le graphe 3D hybride sera construit à partir d'un volume de corrélation sur la scène 3D et des primitives 3D extraites. La surface finale est obtenue par recherche de la coupe de capacité minimale dans ce graphe 3D. La majeure contribution de notre approche consiste à utiliser des primitives 3D extraites et des données externes telles que le réseau routier ou les plans cadastraux pour contraindre le problème d'optimisation et modéliser de manière explicite les occultations et les discontinuités. Le produit final, sera un Modèle Numérique d'Élévation hybride « raster/vecteur », permettant d'exploiter à chaque endroit de la scène, les primitives du niveau le plus élevé que l'on a pu reconstruire de manière fiable.

Mots-clés : images satellitaires, stéréoscopie, modélisation de surfaces 3D, primitives 3D, segmentation, appariement multi-échelles, optimisation à base de flots de graphes, photogrammétrie, Modèle Numérique d'Élévation.

Table des matières

I	Introduction	19
1	Contexte et objectifs	21
1.1	Cadre spécifique : CNES/IGN	21
1.2	La complémentarité entre imageries aérienne et satellitaire	22
1.3	Contexte satellitaire très haute résolution	22
1.3.1	Les satellites PLÉIADES	22
1.4	Les données	25
1.4.1	Simulation des images satellitaires	25
1.4.2	Les Modèles Numériques d'Élévation : MNE	25
1.5	Les objectifs	27
1.6	Plan du manuscrit	28
II	Analyse et stratégie	31
2	État de l'art en modélisation de scènes urbaines	33
2.1	Reconstruction des bâtiments : un problème complexe	33
2.1.1	Complexité des données pour une scène urbaine	33
2.1.2	Problèmes liés à la stéréoscopie	34
2.1.3	Occultations, discontinuités, surfaces à fortes pentes	34
2.2	Les données	35
2.2.1	Principe de la stéréoscopie	35
2.2.2	Appariement stéréoscopique	35
2.2.3	Influence du rapport B/H	37
2.2.4	Qualité a priori d'un MNE	38
2.3	Les principales approches de mise en correspondance	39
2.3.1	Les méthodes surfaciques	40
2.3.2	Les méthodes à base de primitives	42
2.3.3	Les méthodes énergétiques	42
2.4	Les modèles de bâtiments	43
2.4.1	Caricature de surface	44
2.4.2	Approches par modèles paramétriques	44
2.4.3	Approches génériques	44
2.4.4	Bilan	45

3	Stratégie globale	47
3.1	Introduction	47
3.2	Objectif : Modélisation de surfaces 3D	47
3.3	Démarche	47
3.3.1	Reconstruction des primitives	49
3.3.2	Optimisation globale	50
3.3.3	Focalisation	50
III	Reconstruction de primitives 3D	51
4	Reconstruction de segments 3D	53
4.1	Introduction	53
4.2	État de l'art	53
4.3	Algorithme	54
4.3.1	Extraction des segments dans les images	55
4.3.2	Mise en correspondance	55
4.3.3	Reconstruction des segments 3D et validation	56
4.4	Résultats des simulations	57
4.4.1	Simulation des segments 3D	57
4.4.2	Précision par rapport à un modèle de référence	59
4.5	Résultats sur des images réelles	63
4.5.1	Résultats qualitatifs	63
4.5.2	Impact des contraintes	63
4.6	Conclusions	65
4.7	Perspectives	65
5	Reconstruction de facettes 3D par approche segments	67
5.1	Introduction	67
5.2	Méthodologie	67
5.3	Extraction des plans 3D	68
5.4	Délimitation des facettes 3D	69
5.4.1	État de l'art	69
5.4.2	Délimitation des toits par relaxation probabiliste	69
5.4.3	Application à notre problème	70
5.5	Résultats & Évaluations	72
5.5.1	Extraction des demis-plans	72
5.5.2	Délimitation des facettes 3D	76
5.6	Conclusions	77
5.6.1	Avantages	77
5.6.2	Inconvénients	78
5.6.3	Bilan	78
6	Reconstruction de facettes 3D par approche régions	79
6.1	Introduction	79
6.2	État de l'art en reconstruction de surfaces planes	79
6.3	Méthodologie	80
6.4	Segmentation multi-échelles et hiérarchie	81
6.4.1	Modèle de fusion des régions	82

6.4.2	Construction de la hiérarchie	83
6.5	Mise en correspondance des régions	85
6.5.1	État de l'art de la mise en correspondance de primitives	85
6.5.2	Appariement d'une région	87
6.5.3	Reconstruction des plans 3D	92
6.5.4	Qualification des appariements	94
6.5.5	Appariement global des hiérarchies	95
6.6	Résultats	97
6.6.1	Impact de la résolution	101
6.7	Évaluations	103
6.7.1	Qualité de l'appariement	103
6.7.2	Exhaustivité de la reconstruction	104
6.7.3	Temps de calcul	108
6.8	Conclusions	109
6.8.1	Apports	110
6.8.2	Inconvénients	110
6.9	Perspectives	110
6.10	Bilan de la reconstruction de primitives	111
IV	Modélisation de surfaces 3D	113
7	Modélisation de surfaces 3D	115
7.1	Introduction	115
7.2	Minimisation d'énergie : État de l'art	115
7.2.1	Formulation du problème par programmation dynamique	116
7.2.2	Formulation du problème par flot maximal	117
7.2.3	Méthodes coopératives	120
7.2.4	Bilan	121
7.3	Optimisation 3D contrainte	121
7.3.1	Méthodologie	122
7.3.2	Construction du graphe 3D	123
7.3.3	Gestion du graphe 3D hybride	125
7.3.4	Régularisation par surfaces planes : facettes 3D	127
7.3.5	Modélisation des discontinuités : segments 3D	131
7.3.6	Modélisation du sol	133
7.3.7	Résultats	134
7.3.8	Conclusions	137
7.3.9	Perspectives	138
V	Résultats et évaluations	139
8	Plateforme d'évaluation de MNE	141
8.1	Introduction	141
8.2	Évaluation de MNE : État de l'art	141
8.2.1	Qualité intrinsèque	141
8.2.2	Qualité extrinsèque	142
8.3	La plateforme	145

8.3.1	La mesure d'exhaustivité de la reconstruction	145
8.3.2	Évaluation raster-raster	145
8.3.3	Évaluation raster-vecteur	148
8.3.4	Évaluation vecteur-vecteur	148
8.4	Bilan	151
8.5	Perspectives	151
9	Résultats et évaluations	153
9.1	Résultats	153
9.1.1	Amiens 50 cm - $B/H = 0.2$	153
9.1.2	Amiens 70 cm - $B/H = 0.2$	157
9.1.3	Toulouse 60 cm - $B/H = 0.08$	158
9.2	Évaluations	159
9.2.1	Évaluations à 50 cm	159
9.2.2	Influence des paramètres : k et C_f	162
9.2.3	Évaluation à 70 cm	164
9.3	Conclusions	165
10	Conclusions	167
10.1	Analyse méthodologique	167
10.1.1	Contributions	167
10.1.2	Inconvénients	168
10.1.3	Perspectives	169
10.2	Le projet Pléiades	170
10.2.1	Contributions	170
10.2.2	Perspectives	170
10.3	Bilan	171
A	Propagation ascendante de la contrainte épipolaire	173
A.1	Introduction	173
A.2	Algorithme	173
B	Les estimateurs robustes	175
B.1	Introduction	175
B.2	Les M-estimateurs	175
B.2.1	Présentation des M-estimateurs	175
B.2.2	L-estimateurs	178
B.2.3	M-estimateurs généraux	182
B.2.4	Conclusion	191
B.3	Échantillonnage aléatoire	191
B.3.1	LMS : Least Median of Squares	192
B.3.2	Le RANSAC : RANdom Sample Consensus	192
C	Segmentation hiérarchique	195
C.1	Hiérarchies	195
C.1.1	Définitions générales	195
C.1.2	Père et fils, Branches	196
C.1.3	Coupe dans une hiérarchie	196

Table des figures

1.1	Les satellites PLÉIADES-HR	24
1.2	Simulations PLÉIADES - Amiens 50 cm, $B/H = 0.2$. Le MNE est référencé par rapport à la scène topographique et non aux images ce qui explique le changement d'orientation	26
1.3	Les principales étapes du projet PLÉIADES	27
1.4	Objectifs : Entrée-Sortie	28
2.1	Caractéristiques du milieu urbain	34
2.2	Principe de la stéréoscopie	35
2.3	Contrainte épipolaire et altimétrique	36
2.4	Erreur résultant du processus de corrélation	38
2.5	Précision théorique planimétrique et altimétrique	38
2.6	Précision théorique planimétrique et altimétrique	39
2.7	Modèles de bâtiments	45
3.1	MNE hybride raster/vecteur	48
3.2	Stratégie globale	48
4.1	Algorithme de reconstruction de segments 3D	54
4.2	Contrainte épipolaire et altimétrique	55
4.3	Reconstruction et validation de segments 3D	57
4.4	Segment 3D simulé	57
4.5	Précision de reconstruction suivant l'angle du segment par rapport à la base des centres de vue	58
4.6	Précision de reconstruction en fonction du bruit	58
4.7	Précision de reconstruction en fonction du rapport B/H	59
4.8	Modèle de référence pour les simulations	60
4.9	Exhaustivité de la reconstruction en fonction du bruit	60
4.10	Précision en fonction de la bande de tolérance	61
4.11	Simulations - Précision de reconstruction 3D en présence de bruit	62
4.12	Segments 3D reconstruits et leurs projections dans les images	64
5.1	Extraction de facettes 3D : approche segments	67
5.2	Extraction des demi-plans pivotants	68
5.3	Cohérence des demi-plans adjacents	69
5.4	Construction du graphe d'adjacence	71
5.5	Délimitation des facettes 3D par relaxation probabiliste	72
5.6	Extraction de plans 3D - Amiens 50cm	73
5.7	Évaluation des plans par EvalMNE [©]	73

5.8	Evaluations de plans 3D par bâtiment	74
5.9	Exemple centre-ville d'Amiens 50 cm	74
5.10	Evaluations de plans 3D en fonction de la résolution	75
5.11	Exemples de délimitation de facettes 3D	76
5.12	Amiens (France) - 50cm)	77
5.13	Exemple de délimitation des plans 3D, Amiens 50 cm	77
6.1	Schéma global de la reconstruction des facettes 3D par approche région	81
6.2	Construction de la hiérarchie	83
6.3	Différentes échelles de segmentation de la hiérarchie	84
6.4	Segmentation multi-échelles	84
6.5	Mise en correspondance des régions et détection des facettes 2D	87
6.6	Contrainte épipolaire et altimétrique	88
6.7	Application de la contrainte épipolaire par propagation ascendante	89
6.8	Contrainte de recouvrement	91
6.9	Contrainte d'adjacence inter-pyramides pour les régions non appariées	92
6.10	Inconvénients du MNE de corrélation	93
6.11	Calcul du cube de corrélation	94
6.12	Calcul du score d'une facette 3D	95
6.13	Amiens - 50 cm - $B/H = 0.2$	98
6.14	Amiens - 50 cm - $B/H = 0.2$	99
6.15	Toulouse - 60 cm - $B/H = 0.08$	100
6.16	Amiens - 70 cm - $B/H = 0.2$	101
6.17	Amiens - 70 cm - $B/H = 0.2$	102
6.18	Toulouse - 80 cm - $B/H = 0.08$	103
6.19	Procédé d'Évaluation des facettes 2D avec EvalMne, Amiens 50 cm	105
6.20	Impact du seuil de taille à 70 cm	106
6.21	Évaluation 2D en fonction de la résolution	107
6.22	Évaluation 3D des facettes	108
6.23	Évaluations 3D : Impact de la résolution	109
7.1	Exemple d'un graphe 2D orienté, pondéré. Les arêtes de capacité maximale sont illustrées en gras	118
7.2	Construction du graphe 3D : Formulation de Roy et Cox	119
7.3	Schéma global de l'optimisation hybride	122
7.4	Construction du graphe 3D	123
7.5	Capacités de coupe de graphe	124
7.6	Connexion des nappes dans le graphe \mathcal{G}	126
7.7	Problème de surplombs	127
7.8	Exemple idéal d'optimisation hybride avec les facettes 3D	127
7.9	Gestion des nappes en présence de facettes dans le graphe \mathcal{G}	129
7.10	Cas de facettes superposées - Illustration des chemins de coupes en partant de C_1	130
7.11	Illustration du nombre d'arêtes coupées pour une facette	131
7.12	Optimisation globale avec les segments 3D	133
7.13	Optimisation sans primitives	134
7.14	Exemple d'optimisation avec les segments ($K = 0.2, C_f = 0.5$)	135
7.15	Exemple d'optimisation avec les facettes ($K = 0.5, C_f = 0.5$)	136
7.16	Exemple d'optimisation avec le sol	137

8.1	Comparaison ponctuelle entre un MNE et la référence	143
8.2	MNE TRAPU [®] sur Amiens	146
8.3	Exemples de masques créés par EvalMne	146
8.4	Interface de statistiques surfaciques (EvalMne)	147
8.5	Interface de statistiques volumiques (EvalMne)	149
8.6	Distance entre deux plans 3D	150
9.1	Amiens 50 cm - Primitives extraites	153
9.2	Amiens 50 cm, $k = 0.5$ - $C_f = 0$	154
9.3	Amiens 50 cm, $k = 0.2$ - $C_f = 1$	155
9.4	Amiens 50 cm, $k = 0.1$ - $C_f = 0.5$	156
9.5	Amiens 70 cm, $k = 0.5$ - $C_f = 0.5$	157
9.6	Toulouse 60 cm, $k = 0.1$ - $C_f = 0.5$	158
9.7	Toulouse 60 cm, $k = 0.5$ - $C_f = 0.5$	158
9.8	Histogramme des différences d'altitude par classes	159
9.9	images de différence d'altitudes $k = 0.1$ - $C_f = 0.5$	160
9.10	Moyenne des différences d'altitude par classes	161
9.11	Comparaison de profils en présence de facettes ($k = 0.2$ - $C_f = 0.5$)	161
9.12	Influence du coefficient de lissage k - $C_f = 0.5$	162
9.13	Mesure de l' EQM en cm, Robustesse par rapport au coût fixe C_f	163
9.14	Influence de la constance C_f sur le temps de calcul	164
9.15	$k = 0.1$ $C_f = 1.5$	165
B.1	Estimateur L_2 : visualisation	179
B.2	Estimateur L_2 : $res_{\sigma=1}(x)$ sur les intervalles $[-2,2]$ et $[-2.5,22.5]$	180
B.3	Estimateur L_1 : visualisation	180
B.4	Estimateur L_1 : $res_{\sigma=1}(x)$ sur les intervalles $[-2,2]$ et $[-2.5,22.5]$	181
B.5	Estimateur L_p : visualisation	182
B.6	Estimateur L_p : $res_{\sigma=1}(x)$ sur les intervalles $[-2,2]$ et $[-2.5,22.5]$	182
B.7	Estimateur $L_1 - L_2$: visualisation	184
B.8	Estimateur $L_1 - L_2$: $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2,2]$	184
B.9	Estimateur $L_1 - L_2$: $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2.5,22.5]$	184
B.10	Estimateur Fair : visualisation	185
B.11	Estimateur Fair : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2,2]$	185
B.12	Estimateur Fair : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2.5,22.5]$	186
B.13	Estimateur Huber : visualisation	186
B.14	Estimateur Huber : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2,2]$	187
B.15	Estimateur Huber : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2.5,22.5]$	187
B.16	Estimateur Cauchy : visualisation	187
B.17	Estimateur Cauchy : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2,2]$	188
B.18	Estimateur Cauchy : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2.5,22.5]$	188
B.19	Estimateur Tukey : visualisation	189
B.20	Estimateur Tukey : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2,2]$	189
B.21	Estimateur Tukey : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2.5,22.5]$	189
B.22	Estimateur Welsch : visualisation	190
B.23	Estimateur Welsch : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2,2]$	190
B.24	Estimateur Welsch : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2.5,22.5]$	191
C.1	Représentation d'une hiérarchie sous forme de dendrogramme	195

Liste des tableaux

1.1	Les principales caractéristiques des satellites Haute Résolution	23
2.1	Précisions théoriques sur les simulations PLÉIADES	39
4.1	Impact des contraintes pour la détection des segments 3D	63
6.1	Contraintes d'appariement des régions	103
6.2	Évaluation de la coupe optimale	104
6.3	Influence de la taille maximale de région	105
6.4	Influence de la taille minimale de région	106
7.1	Capacités des arêtes du graphe	131
B.1	Définition de l'estimateur L_2	179
B.2	Définition de l'estimateur L_1	180
B.3	Définition de l'estimateur L_p	181
B.4	Définition de l'estimateur $L_1 - L_2$	183
B.5	Définition de l'estimateur Fair	185
B.6	Définition de l'estimateur Huber	186
B.7	Définition de l'estimateur Cauchy	187
B.8	Définition de l'estimateur Tukey	188
B.9	Définition de l'estimateur Welsch	190

Première partie

Introduction

Chapitre 1

Contexte et objectifs

1.1 Cadre spécifique : CNES/IGN

Cette thèse a été réalisée dans le cadre d'une coopération tripartite IGN - CNES - Université René Descartes, Paris 5. Cette étude a été lancée suite à l'apparition des premiers satellites américains submétriques. La communauté de l'imagerie aérienne et spatiale est en plein bouleversement suite à une progression rapide des technologies et le renforcement de la concurrence dans un contexte international. La commercialisation des données satellitaires de résolution submétrique ouvre de nouvelles perspectives et de nouveaux champs d'applications. Ainsi le CNES (Centre National d'Études Spatiales), riche de son expérience des filières SPOT et Hélios travaille activement sur le concept des satellites PLÉIADES. Cette thèse a été menée par le CNES pour évaluer la qualité de la reconstruction de surfaces 3D à partir des images des futurs satellites PLÉIADES, successeurs de SPOT5.

Le deuxième partenaire est l'Institut Géographique National (IGN), traditionnellement fournisseur de cartes et de bases de données géographiques, qui s'est intéressé aux méthodes de production de ces bases de données tridimensionnelles. La première chaîne de production TRAPU[®], entièrement manuelle et basée sur des techniques de saisie photogrammétrique traditionnelle sur des appareils analogiques a été mise au point dans les années 80. Cependant, la réalisation manuelle de telles bases de données constitue un processus très délicat et coûteux ce qui a conduit à l'abandon de cette chaîne de production. Depuis lors, l'IGN a relancé plusieurs recherches sur l'automatisation des processus de production de bases tridimensionnelles. Ces recherches sont menées au sein du laboratoire MATIS (Méthodes d'Analyse et de Traitement des Images pour la Stéréorestitution) et se focalisent plus particulièrement sur la reconstruction du bâti et s'articulent au sein d'une action de recherche locale très active : BRIGE (Bâtiments Reconstitués sur Images à Grande Échelle). Les images utilisées sont des images aériennes numériques à très grande échelle (10 cm - 25 cm).

Enfin, l'encadrement académique a été assuré au niveau de l'université René Descartes par l'équipe SIP (Systèmes d'Information pour la Perception) du laboratoire CRIP5. Un axe de recherche du laboratoire porte sur la reconstruction 3D et un projet est également réalisé en collaboration avec le CNES sur des problématiques de détection d'objets cartographiques (routes, carrefours et ronds-points) à partir d'images satellitaires.

On présentera dans la suite de ce chapitre les caractéristiques du contexte satellitaire très haute résolution et ses principaux enjeux. On s'intéressera tout particulièrement au projet des satellites PLÉIADES. Les objectifs fixés pour cette thèse seront resitués par rapport à ce projet global. On conclura ce chapitre par le plan du manuscrit.

1.2 La complémentarité entre imageries aérienne et satellitaire

Les satellites de résolution métriques sont apparus suite aux développements technologiques et aux besoins stratégiques croissants. Pour des applications militaires, les paramètres privilégiés sont la fréquence de revisite, la rapidité d'accès, etc. Par contre, les opérations civiles privilégient plutôt la taille des zones de couverture et la qualité des images. Face à ces exigences, les images aériennes présentent des avantages par rapport aux images satellitaires :

- une meilleure adaptation aux aléas météorologiques, l'avion pouvant exploiter une période de beau temps dans sa totalité pour couvrir une grande zone.
- une meilleure capacité d'acquisition multi-vues (recouvrement longitudinal de 80% et latéral de 60% pour les caméras matricielles) qui est nécessaire pour assurer une bonne précision stéréoscopique surtout en milieu urbain dense.
- une capacité à fournir des images acquises à basse altitude, donc moins dégradées par la diffusion atmosphérique.

Les principaux inconvénients des prises de vue spatiales sont :

- le manque de souplesse pour les caractéristiques des images. Celles-ci devant être fixées au lancement du satellite.
- la rareté des données satellitaires rend également difficile la couverture de grandes zones.

Néanmoins, le satellite reste plus performant pour certaines applications nécessitant :

- une grande réactivité : situations d'urgences, applications liées aux risques naturels.
- une répétitivité : problématique de suivi de phénomènes évolutifs.
- un accès éloigné (applications dans le domaine des risques) ou bien sur des sites où il est difficile d'obtenir un plan de vol.

Le principal atout des données spatiales à haute résolution est l'accessibilité des lieux. La possibilité de couvrir potentiellement toute la surface du globe avec des données à haute résolution, permet aux sociétés d'ingénierie de disposer de cartes actualisées sur leur zone d'intérêt, dans les domaines du génie civil, des télécommunications, etc. C'est le premier secteur de vente des images Ikonos (cf. tableau 1.1).

1.3 Contexte satellitaire très haute résolution


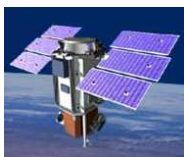


Des images de même nature, mais d'origines différentes, aériennes ou satellitaires sont maintenant disponibles à l'utilisateur. Le tableau 1.1 récapitule l'ensemble des satellites à haute et moyenne résolution et permet de faire un comparatif de quelques caractéristiques [DKB⁺03].

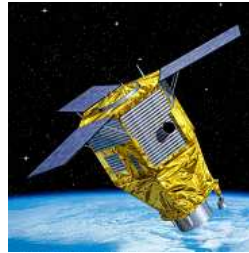
Si la résolution spatiale est le premier critère de classement intuitif, d'autres paramètres sont également à prendre en compte par l'utilisateur tel que la répétitivité du passage pour un lieu donné, la largeur de la zone couverte, l'archivage et l'accessibilité des données, etc. Pour cette raison, ce récapitulatif comporte les satellites à haute et moyenne résolution, entre 1 et 2 mètres. Dans la section suivante, on s'intéressera tout particulièrement aux satellites PLÉIADES.

1.3.1 Les satellites PLÉIADES

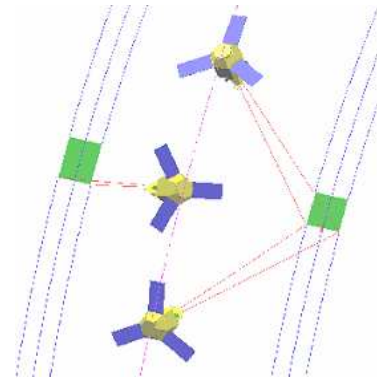
Le projet des satellites PLÉIADES [Plé05], est conçu autour de plusieurs constellations de petits satellites mono-instruments (cf. figure 1.1-a) et d'une architecture d'ensemble évolutive en fonction du déploiement des composantes spatiales (cf. figure 1.1-b).

TAB. 1.1 – Les principales caractéristiques des satellites Haute Résolution

Satellite	Caractéristiques du satellite		
Ikonos 2	http://www.spaceimaging.com/products/ikonos/index.htm		
 <p>Le satellite Ikonos 2 Lancement 09/1999</p>	Domaine spectral	Panchromatique	multispectral
	Résolution	1 mètre	4 mètres
	Nombre de canaux	1	4
	bandes spectrales	450-900 nm	450-520 nm
			520-600 nm
			625-695 nm
			760-900 nm
	Fauchée	11 km	
	Surface Image	11 km * 11 km à 1000 km	
	Altitude d'orbite	680 km	
Dynamique	11 bits/canal		
QuickBird 2	http://www.digitalglobe.com/		
 <p>Le satellite QuickBird 2 Lancement 18/10/2001</p>	Domaine spectral	Panchromatique	multispectral
	Résolution	0.61 mètre	2.44 mètres
	Nombre de canaux	1	4
	bandes spectrales	450-900 nm	450-520 nm
			520-600 nm
			630-690 nm
			760-890 nm
	Fauchée	16.5 km	
	Surface Image	17 km * 17 km, bande de 16.5 à 165 km	
	Altitude d'orbite	450 km	
Dynamique	11 bits/canal		
Orbview 3	http://www.orbimage.com/		
 <p>Le satellite Orbview 3 Lancement 26/06/2003</p>	Domaine spectral	Panchromatique	multispectral
	Résolution	1 mètre	4 mètres
	Nombre de canaux	1	4
	bandes spectrales	450-900 nm	450-520 nm
			520-600 nm
			625-695 nm
			760-900 nm
	Fauchée	8 km	
	Surface Image	8 km * 8 km	
	Altitude d'orbite	470 km	
Dynamique	11 bits/canal		
Eros A1	http://www.imagesatintl.com/		
 <p>Le satellite Eros A1 Lancement 5/12/2000</p>	Domaine spectral	Panchromatique	multispectral
	Résolution	1.8 mètre	
	Nombre de canaux	1	
	bandes spectrales	500-900 nm	
	Fauchée	13.5 km	
	Surface Image	13.5 km * 13.5 km à 40km	
	Altitude d'orbite	480 km	
	Dynamique	11 bits/canal	



(a) Satellites Pléiades



(b) Acquisitions simultanées stéréo et tri-stéréo

FIG. 1.1 – Les satellites PLÉIADES-HR

1.3.1.1 Les caractéristiques optiques

Le système PLÉIADES-HR devrait comprendre deux satellites placés à 180° sur une même orbite, dont les caractéristiques prévues sont :

- orbite héliosynchrone à 695 km d'altitude.
- une capacité d'accès journalière en tout point du globe (avec dépointage maximal à 47°).
- une résolution de 0,7 m au nadir en mode panchromatique, 2.8 m en mode multispectral (B0-bleu, B1-vert, B2-rouge, B3-PIR).
- image codée en 12 bits.
- une fauchée de prise de vue de 20 km au nadir
- une capacité d'acquérir dans un même passage une mosaïque d'images d'une taille de 120 km par 120 km.
- une capacité d'acquisition quasi instantanée de couples stéréoscopiques de 20 km par 350 km, avec une capacité de tri-stéréoscopie incluant une visée au nadir, particulièrement utile pour les applications en milieu urbain.
- une capacité de couverture d'images sans nuages de 2500000 km^2 par an.
- localisation des images ($< 1\text{m}$ avec des points d'appui). Sans prise de points d'appui, une précision inférieure à 10 m (dans 90% des cas) est prévue. La distorsion interne de l'image devrait être telle que l'erreur de localisation à partir de points d'appui, et en utilisant une modélisation linéaire de l'erreur d'altitude, reste inférieure à 0.9 pixel sur une période temporelle de 12 secondes.

Concernant la capacité d'acquisition du système PLÉIADES-HR (cf.figure 1.1-b), le corps entier du véhicule est orienté (roue à inertie) pour pointer la ligne de visée. Les satellites PLÉIADES-HR permettront l'acquisition d'images dans n'importe quelle direction à l'intérieur d'un cône de 30° centré sur le nadir (et jusqu'à 47° avec une qualité d'image moindre). Cette capacité permettra d'acquérir des couples et même des triplets d'images simultanées de résolution 70 cm (nadir) selon des rapports B/H variés (0.15 à 0.8). B étant la base stéréoscopique et H l'altitude des points de vue par rapport au terrain. La définition du rapport B/H et son influence sur la précision théorique de la reconstruction 3D seront détaillées dans la section 2.2.3.

1.3.1.2 Les principales applications

Le satellite PLÉIADES est un système multi-capteurs destiné à satisfaire le plus grand nombre d'utilisateurs. Les applications envisageables sont donc très variées ; chacune requiert des caractéristiques spécifiques, liées non seulement à la nature des informations recherchées, mais également à ses particularités géographiques, climatiques, socio-économiques, etc. Les applications les plus représentatives ont été identifiées et regroupées par domaines :

- Cartographie : cartographie de référence, aménagement, urbanisme, télécommunications ;
- Agriculture : agriculture de précision, contrôle, statistiques ;
- Foresterie : production, protection (Feux de forêts), écologie ;
- Risques : les risques hydrologiques (inondations), la géologie dynamique (risques sismiques et volcaniques, risques de glissements de terrain), Prospection géologique ;
- Applications marines : océans, glaces de mer, littoral ;
- Défense : applications militaires et de sécurité.

1.4 Les données

Afin de définir et développer les outils nécessaires à l'exploitation opérationnelle des futures images sub-métriques optiques (aspects tridimensionnels, reconnaissance de formes, analyse de texture, détection de changements, etc.), les études doivent s'appuyer sur des données simulées (acquises à partir de campagnes aéroportées) utilisées de manière communautaire sur un ensemble de sites pilotes. La validation des produits et services sera faite avec ces données simulées.

Dans le cadre de cette thèse, on dispose en entrée de simulations PLÉIADES à des résolutions de 50, 60 et 70 cm avec différents B/H variant entre [0.08-0.2]. On exclut la multiscopie dans le cadre de cette étude. On travaillera avec des couples stéréoscopiques panchromatiques présentant un recouvrement de 60%.

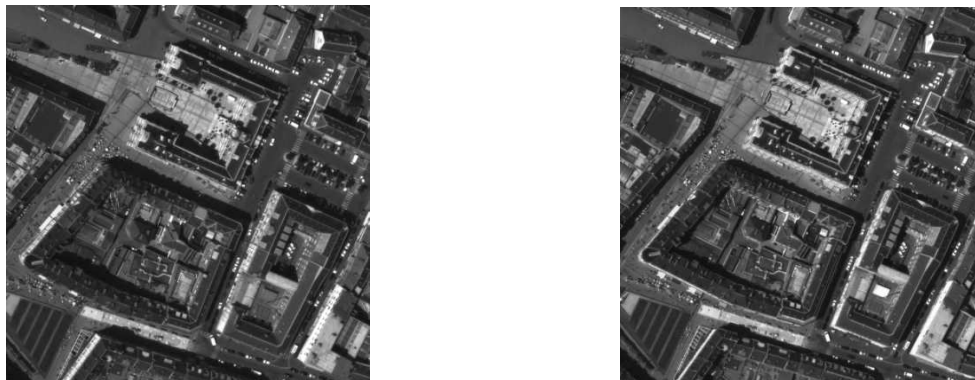
1.4.1 Simulation des images satellitaires

Dans le cadre de cette thèse, les images des futurs satellites PLÉIADES n'étant pas encore disponibles, les images de référence seront simulées à partir des images aériennes prises depuis un avion. Le processus de simulation consiste à dégrader la qualité radiométrique des images aériennes en tenant compte des caractéristiques des capteurs du satellite. Pour les expérimentations, la ville de Toulouse a été couverte avec des images aériennes brutes à 26 cm permettant de générer des simulations PLÉIADES à 60, 80 et 100 cm de résolution avec un rapport B/H de 0.08. La ville d'Amiens a été également couverte par des images aériennes de 25 cm, permettant de simuler des images satellitaires à 50 et 70 cm avec un rapport $B/H = 0.2$ (cf. figure 1.2(a)). Ces simulations PLÉIADES ont été réalisées par le CNES/QTIS.¹ Ces images sont géoréférencées, c'est à dire que l'on dispose de tous les paramètres de prise de vue et donc des paramètres de projection.

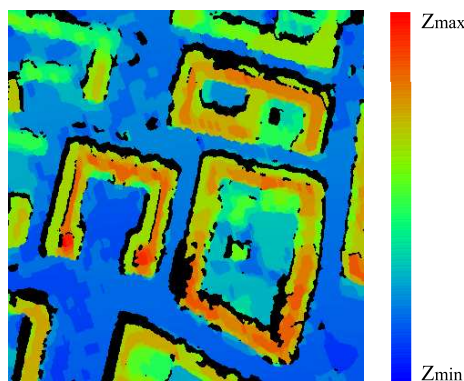
1.4.2 Les Modèles Numériques d'Élévation : MNE

On dispose en entrée de toute information directement et automatiquement issue des images simulées. C'est le cas, en particulier, du MNE (Modèle Numérique d'Élévation), obtenu par des techniques de corrélation [Bai97] à partir du couple stéréoscopique (cf. figure 1.2(b)).

¹CNES : Centre National des Études Spatiales, Toulouse
QTIS : Qualité et Traitement des Images Spatiales.



(a) Couple stéréoscopique



(b) MNE obtenu par des techniques de corrélation

FIG. 1.2 – Simulations PLÉIADES - Amiens 50 cm, $B/H = 0.2$. Le MNE est référencé par rapport à la scène topographique et non aux images ce qui explique le changement d'orientation

Un MNE est une représentation maillée de toute la surface topographique incluant sol, bâtiments, végétation, etc. Cette représentation est souvent en $2D^{1/2}$ puisqu'il s'agit d'une grille régulière qui renvoie pour chaque point (x, y) , son altitude terrain. Le géoréférencement de cette grille permet d'avoir les coordonnées 3D correspondantes du point (X, Y, Z) .

Beaucoup de travaux ont été réalisés pour la génération de MNE. On ne détaillera pas ce problème ici. Le MNE utilisé dans notre cas, est celui développé par C. Baillard [Bai97]. Il est basé sur la technique de programmation dynamique le long des lignes épipolaires et prend en compte les problèmes de parties cachées. Par conséquent, le MNE final n'est pas dense, les parties cachées non appariées sont affectées d'une valeur par défaut. Ce MNE est également pixellaire, on verra par la suite les limitations immédiates en cas d'application dans un contexte satellitaire. Il n'est pas précis mais robuste et a un bon comportement au niveau des discontinuités. Par conséquent, ce MNE sera essentiellement utilisé pour définir une bande de tolérance robuste qui fournit des altitudes minimales et maximales fiables pour chaque pixel de la zone étudiée.

1.5 Les objectifs

Cette thèse se situe dans le cadre d'un projet global d'interprétation de scènes urbaines et de reconstruction de bâtiments à partir d'images satellitaires très haute résolution. Ce projet constitue une étude prospective des capacités des futurs satellites PLÉIADES-HR à reconstruire le relief des paysages urbains. Une évaluation de la précision de reconstruction permettra de développer les outils nécessaires à l'exploitation opérationnelle des images Pléiades. La figure 1.3 illustre les principales étapes du projet et son avancement. Les premières études ont commencé sur la composante capteur des satellites. Les caractéristiques des images ont été fixées au courant de l'année 2001. Le lancement des satellites PLÉIADES-HR est prévu pour fin 2008.

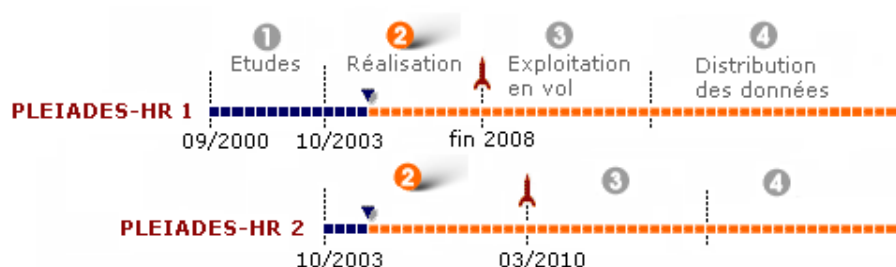


FIG. 1.3 – Les principales étapes du projet PLÉIADES

L'objectif final de ce projet sera la reconstruction automatique de modèles 3D de bâtiments à partir d'un couple stéréoscopique ou bien d'un triplet d'images de simulation PLÉIADES en utilisant éventuellement des données externes comme le réseau routier ou des plans cadastraux.

Dans le cadre de cette thèse, la multiscopie est exclue et on ne traitera que le cas de la stéréoscopie simple. On dispose en entrée de couples stéréoscopiques panchromatiques à [50-70 cm] de résolution. La résolution définitive des capteurs a été fixée au courant de cette thèse et sera finalement de 70 cm au nadir. En outre, on s'intéresse dans le cadre de cette thèse uniquement aux faibles valeurs de B/H [0.05-0.02], sachant que les satellites PLÉIADES pourront fournir des images avec des rapports B/H allant jusqu'à 0.8. (cf. section 1.3.1.1).

On s'autorise également à utiliser les données directement et automatiquement obtenues à partir des images, tel est le cas du modèle numérique d'élévation (cf. section 1.4.2). À partir de ces données sources, notre objectif sera de proposer un système automatique de modélisation de surfaces 3D de scènes urbaines en contexte satellitaire (cf. figure 1.4). En outre, on se place dans un contexte urbain. On ne s'intéressera qu'à la reconstruction des structures de toits, seuls éléments de bâtiments visibles depuis les images satellitaires. Pour avoir une certaine pérennité, le système devra facilement pouvoir intégrer des données nouvelles, telles que les limites cadastrales ou encore le réseau routier 3D qui permettraient de guider l'extraction de primitives.

Toutefois, le contexte satellitaire très haute résolution présente des limitations qui sont d'ores et déjà prévisibles. Tout d'abord la résolution des images, certes élevée (50-70 cm) n'est pas comparable aux résolutions des images aériennes (8cm-25 cm) et ne permet pas de voir tous les détails présents sur les toits. Le contexte stéréoscopique présente également de nombreuses difficultés compte tenu des parties cachées, des zones occultées par la végétation ou les ombres portées. Et finalement, le faible rapport B/H a l'avantage de réduire les parties cachées certes et les distorsions géométriques entre images, mais la précision de reconstruction altimétrique est dégradée. Les algorithmes proposés devront composer avec la très faible base stéréoscopique.

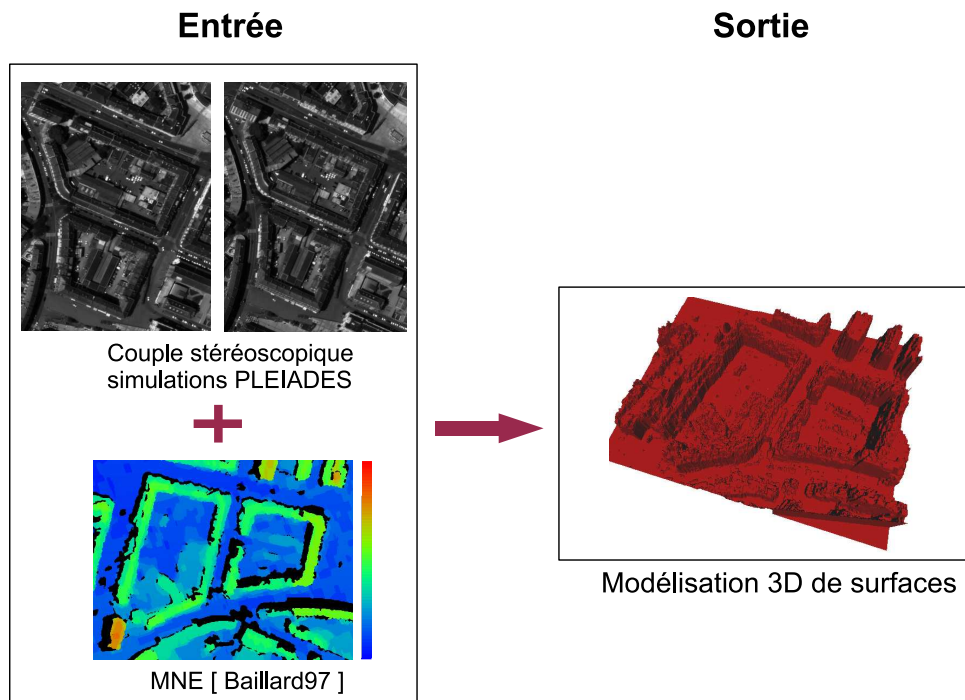


FIG. 1.4 – Objectifs : Entrée-Sortie

En résumé, compte tenu des limitations du contexte, on ne cherchera pas à reconstruire des modèles polyédriques 3D des bâtiments en gardant la topologie. Notre objectif sera de proposer un système **automatique** pour la modélisation 3D de la scène, en proposant un relief simplifié regroupant les bâtiments, les principaux toits et les routes. La description de la scène devra être **dense** pour des applications telles que le calcul d'orthophotos, ou les survols virtuels. Le système doit également être **ouvert** et facilement **extensible** pour intégrer des nouvelles données. Une attention particulière sera apportée à l'évaluation de la modélisation 3D de surfaces à partir d'un couple stéréoscopique (précision altimétrique, exhaustivité, robustesse, etc.).

Cette étude permettra de déterminer les capacités d'utilisation des futurs satellites PLÉIADES et de la commercialisation des images pour des applications cartographiques. Les questions qui se posent actuellement : Quels produits pourra-t-on fournir à partir de ces images ? Quel niveau de détail peut-on avoir en reconstruction de bâtiments ? Quel est l'impact de la résolution, de la qualité de l'image sur la précision de reconstruction ? Est-il possible d'adapter, à moindre coût, des méthodes utilisées en contexte aérien à un contexte satellitaire ? Quelles sont les principales limitations ? Quelles sont les orientations à prendre ? Un certain nombre de questions auxquelles il faudra répondre tout au long de ce travail.

1.6 Plan du manuscrit

Suite à la présentation du contexte satellitaire et des objectifs, quatre grandes parties se détachent dans le manuscrit de la thèse. Une première partie "**Analyse et Stratégie**" présente, après un rappel sur le principe de stéréoscopie et l'influence du rapport B/H du couple stéréoscopique (cf. chapitre 2), un état de l'art détaillé sur les techniques de mise en correspon-

dance d'images ainsi que les modèles et les stratégies de reconstruction de bâtiments. Suite à un bilan de l'état de l'art, la stratégie choisie sera développée et justifiée dans le chapitre 3. Dans la deuxième partie "**Reconstruction de primitives**", les algorithmes de reconstruction de segments 3D et les deux approches de reconstruction de facettes 3D sont détaillés. Les résultats et les évaluations seront présentés à la fin de chaque chapitre, pour chaque approche de reconstruction. La troisième partie "**Modélisation de surfaces 3D**" propose tout d'abord un état de l'art détaillé sur les approches de reconstruction de surfaces par minimisation d'énergie (cf. section 7.2) ensuite, le processus d'optimisation contrainte pour la modélisation 3D de la scène sera détaillé (cf. section 7.3). Dans la dernière partie "**Résultats et Évaluations**", on présentera tout d'abord la plateforme d'évaluation développée durant cette thèse, qui permet de réaliser des évaluations aussi bien au niveau raster qu'au niveau vecteur, ainsi que les résultats obtenus sur différentes zones urbaines et leurs évaluations en utilisant la plateforme.

Deuxième partie

Analyse et stratégie

Chapitre 2

État de l'art en modélisation de scènes urbaines

Introduction

Le cadre de cette thèse porte de manière plus générale sur la modélisation de scènes urbaines à partir d'un couple d'images satellitaires. Cette modélisation comporte plusieurs niveaux de description de la scène allant de la mise en correspondance des images, la reconstruction de surfaces 3D à la reconstruction de modèles 3D de bâtiments. Dans ce chapitre, on pose tout d'abord la problématique de la reconstruction 3D des bâtiments et les difficultés rencontrées en milieu urbain dense. On rappelle ensuite le principe de stéréoscopie et l'influence du rapport B/H sur la précision de la reconstruction 3D. Un état de l'art sur les différentes approches de mise en correspondance d'images est présenté. Les méthodes permettent d'aboutir à des ensembles de points 3D non structurés ou structurés en utilisant les primitives. Cette mise en correspondance permet de générer, à un plus haut niveau, différents modèles 3D de bâtiments qui seront également détaillés dans ce chapitre.

2.1 Reconstruction des bâtiments : un problème complexe

2.1.1 Complexité des données pour une scène urbaine

Étant donnée la haute résolution des images [50-70 cm], plusieurs détails de la scène urbaine sont présents tels que : les cheminées, les chiens-assis, les voitures, les passages piétons, etc. Leur interprétation directe est difficile et il est souvent nécessaire de les structurer en données plus compactes afin de mieux interpréter la scène. Certains de ces objets constitueront du bruit selon ce que l'on recherche. La complexité de la scène peut être séparée en deux parties, une complexité "sémantique" et une complexité "signal". La complexité "signal" est reliée directement aux images sources et dépend :

- **du nombre d'images** qui permet d'assurer la couverture totale et la reconstruction de la ville (monoscopie, stéréoscopie, multi- vues).
- **du type d'images** : images N/B, couleur, multispectrale, capteurs laser, etc.

La complexité "sémantique " est plus reliée à l'interprétation de la scène et dépend à la fois :

- **du nombre d'objets** qui composent la scène (bâtiments, arbres, voitures, ...). Le nombre d'objets dépend de la nature du milieu (zone rurale, périurbaine ou urbaine dense) et des conditions d'observations (résolution, qualité des images, saisons...).

- **de la diversité des objets** qui composent la scène : arbres, arbustes, immeubles, maisons, haies, passages piétons, voitures (cf. figure 2.1-a), etc. Ces objets présentent des caractéristiques de forme, de taille, de radiométrie différentes et nécessitent par conséquent des traitements spécifiques à chaque type d'objet.
- **de la complexité des objets** qui composent la scène. Les bâtiments présentent des formes différentes et il est difficile d'avoir une reconstruction générique pour tous les bâtiments. À ce niveau, se pose la question du niveau de reconstruction, quel détail cherche-t-on à reconstruire ? Le niveau d'interprétation dépend évidemment de la résolution de l'image, mais aussi de l'application visée.

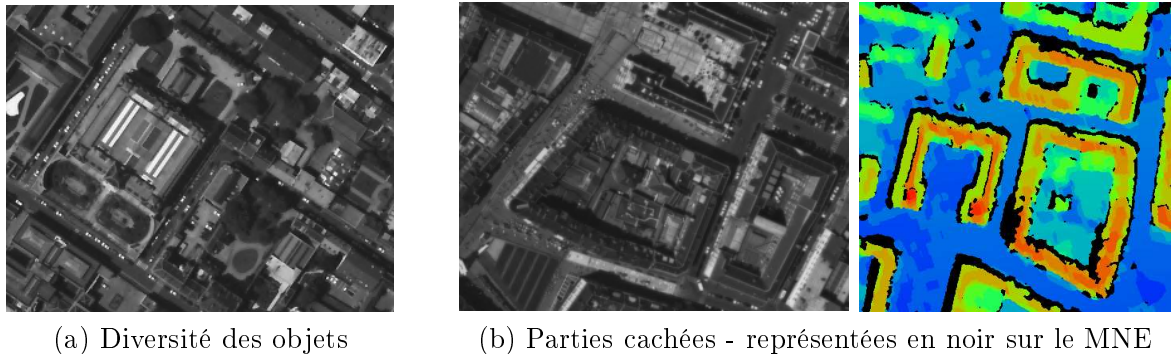


FIG. 2.1 – Caractéristiques du milieu urbain

2.1.2 Problèmes liés à la stéréoscopie

D'autres problèmes se rajoutent dans le cas de couples d'images stéréoscopiques. Le problème de mise en correspondance devient alors complexe. Les éléments les plus fréquents sont présentés dans ce qui suit.

- **Radiométrie** : Tout d'abord, la luminance de prise de vue entre deux images peut être différente surtout si les deux images ne sont pas prises en même temps. De plus, les surfaces éclairées des objets ne réfléchissent pas uniformément la lumière si elles ne sont pas lambertiennes et ceci pose des problèmes pour la mise en correspondance.
- **Le temps** : Les images peuvent ne pas être prises en même temps. Les conditions d'éclairage ainsi que l'orientation des ombres peuvent changer. Des objets de la scène peuvent être modifiés ou s'être déplacés tels que les véhicules.
- **Occultations** : Le principal problème de mise en correspondance en stéréoscopie provient des occultations. Les régions d'occultation correspondent à des régions qui sont vues dans une image et cachées dans l'autre, typiquement les façades de bâtiments (cf. figure 2.1-b). Dans ce cas, les régions ne peuvent être appariées et aucune profondeur ne leur est attribuée (régions illustrées en noir sur le MNE).

2.1.3 Occultations, discontinuités, surfaces à fortes pentes

Les scènes en milieu urbain dense sont composées de différents objets constituant le sursol tels que les arbres et les bâtiments. Trois principales difficultés se présentent alors. Tout d'abord, les objets du sursol vont créer des occultations au niveau du sol par les ombres portées, mais aussi

entre eux dans le cas de milieu urbain très dense (végétation avec les bâtiments, les bâtiments entre eux). Ces zones d'occultations compliquent le problème.

De plus, la plupart des visées sont quasi verticales donc les façades constituent des surfaces à fortes pentes qui ne pourront pas être appariées directement. En pratique, ces façades vont correspondre à des régions trop déformées d'une image à l'autre. Le problème des façades cachées se pose aussi et rend leur appariement impossible (cf. figure 2.1-b).

Le troisième problème se pose au niveau des discontinuités au sein des bâtiments qui posent également un problème lors de la mise en correspondance.

2.2 Les données

2.2.1 Principe de la stéréoscopie

La figure 2.2 rappelle quelques éléments intervenant en stéréorestitution. Cette figure représente les deux centres ou sommets de prise de vue S_1 et S_2 , à partir desquels ont été acquises les images I et J . Les deux emprises au sol des images présentent une zone de recouvrement où s'applique justement le principe de stéréoscopie. Si un point M sur le terrain est vu dans les images respectivement en m_1 et m_2 , il est possible en calculant l'intersection des rayons perspectifs S_1m_1 et S_2m_2 de retrouver la position 3D du point M .

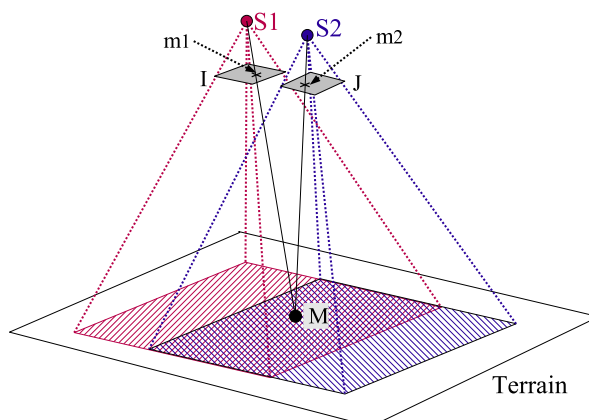


FIG. 2.2 – Principe de la stéréoscopie

Ce calcul nécessite toutefois de connaître la position des sommets de prise de vue S_1 et S_2 , ainsi que les orientations des caméras (roulis, tangage, lacet). Ces éléments, qui constituent la calibration externe des caméras, sont calculés par aérotriangulation [KE01a]. Cet aspect technique ne sera pas explicité dans ce manuscrit. Il faut également connaître la calibration interne de la caméra, mais on ne s'intéresse pas à ce problème dans ce manuscrit et on supposera que ces éléments sont parfaitement connus. Le problème qui se pose est, comment déterminer de manière automatique, fiable et précise, les positions des points homologues m_1 et m_2 ? Ce problème d'appariement constitue **la mise en correspondance**, dont l'état de l'art est détaillé dans la section 2.3.

2.2.2 Appariement stéréoscopique

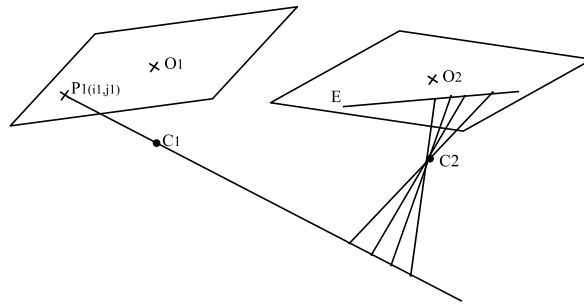
Cette section détaille les contraintes généralement utilisées pour l'appariement de pixels en stéréoscopie.

2.2.2.1 La contrainte épipolaire

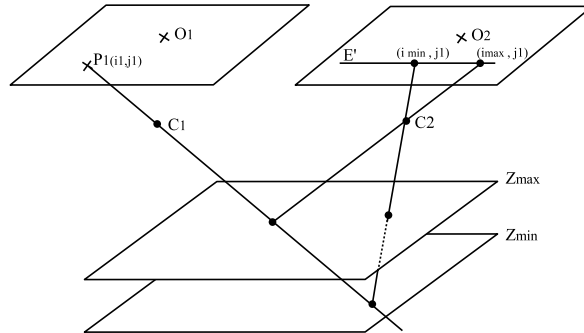
L'appariement dans l'espace image nécessite l'orientation relative des images. Supposons un point P_1 dans l'image de référence, on cherche à retrouver son homologue P_2 dans la seconde image (cf. figure 2.3(a)). La position de P_2 est géométriquement contrainte. Tous les appariements possibles dans la seconde image sont sur une ligne E appelée ligne épipolaire qui est la projection dans le plan de l'image du plan 3D passant par (P_1, C_1, C_2) . Le problème d'appariement en 2D est ainsi réduit à une mise en correspondance 1D le long de la ligne épipolaire. Cette contrainte constitue la **contrainte épipolaire**.

2.2.2.2 Définition d'un intervalle pour l'espace de recherche

Pour la mise en correspondance des primitives dans l'espace image, l'espace de recherche peut être réduit en utilisant les altitudes minimale et maximale pour chaque pixel (cf. figure 2.3(b)). Ces altitudes peuvent être obtenues à partir du MNE par exemple.



(a) Appariement stéréoscopique depuis l'espace image



(b) Définition d'un intervalle de l'espace de recherche

FIG. 2.3 – Contrainte épipolaire et altimétrique

2.2.2.3 Définition de la notion de disparité/profondeur

On dit que deux pixels p_1 et p_2 sont **homologues** ou **conjugués** s'ils correspondent aux projections images du même point physique P . On définit alors la **disparité** (appelée encore **parallaxe**), $disp(p_1, p_2)$, la variation de position entre les pixels p_1 et p_2 d'une image à une autre.

La géométrie de prise de vue permet d'interpréter la valeur de disparité comme une **profondeur** (distance du point P au système de prise de vue), liée elle-même à l'altitude du point P .

2.2.2.4 Les contraintes générales

La plupart des algorithmes de stéréovision utilisent des contraintes fortes sur la nature et la structure de scène dont les plus courantes sont détaillées dans ce qui suit :

- **Contrainte lambertienne** : Pour que les images d'un point 3D aient la même intensité, il faut que la surface soit lambertienne c'est-à-dire que la luminosité réfléchie par la surface ne dépend pas de la position d'observation. Le modèle de réflectance lambertien est le plus simple. Les modèles mathématiques modélisant la réflexion deviennent rapidement complexes étant donnée la diversité des objets. Pour compenser l'erreur réalisée par cette approximation, les algorithmes de mise en correspondance se basent sur des éléments peu sensibles à cette contrainte tels que les contours ou les segments.
- **Contrainte d'unicité** : Elle consiste à imposer un seul appariement par primitive. Un pixel d'une image ne peut avoir qu'un seul point homologue dans la seconde image.
- **Contrainte de réciprocité** : Impose un traitement symétrique des deux images, seuls sont retenus les appariements réciproques entre les deux images.
- **Contrainte d'ordre** : Si un objet A est à droite d'un objet B dans une image, son homologue A' dans la seconde image sera à droite de l'homologue de B , B' . Cette contrainte n'est pas vérifiée pour les objets transparents par exemple.
- **Contrainte fronto-parallèle** : On suppose que la surface observée est parallèle au plan rétinien des deux caméras. Cette contrainte permet d'assurer que les voisinages autour des projections d'un point 3D sont identiques dans les deux images.

2.2.3 Influence du rapport B/H

Les deux paramètres importants de la prise de vue stéréoscopique sont la base de prise de vue B qui est la distance des centres de prise de vue et la hauteur de vol H (cf. figure 2.4).

La précision 3D est affectée par les erreurs d'aérottriangulation et les erreurs de corrélation. Ses composantes planimétrique et altimétrique dépendent des angles d'observation et du rapport base sur hauteur (B/H) (cf. figure 2.4). Supposons que l'orientation relative soit parfaite, on s'affranchit ainsi des erreurs dues à l'aérottriangulation. On peut estimer théoriquement la précision intrinsèque du processus de corrélation. Notons e_{corr} , l'erreur de corrélation exprimée en pixels et r_o la résolution du pixel sol exprimée en mètres.

L'erreur altimétrique est donnée par la relation suivante [KE01b] :

$$e_{alti} = \frac{H}{B} \cdot r_o \cdot e_{corr} \text{ et } \sigma_{alti} = \frac{H}{B} \cdot r_o \cdot \sigma_{corr} \quad (2.1)$$

L'erreur planimétrique est radiale et dépend donc de la position de (i_1, j_1) dans l'image : $e_{plani} = tg(\alpha) \cdot e_{alti} = \| \vec{O_1 P_1} \| \cdot e_{alti} / f$ où f est la distance focale exprimée en pixels.

D'après l'équation 2.1, un fort rapport B/H , permet d'avoir une bonne précision altimétrique par contre les distorsions géométriques et les parties cachées entre les images augmentent. Réciproquement, un faible B/H , rend l'appariement plus fiable, mais la précision altimétrique de reconstruction est dégradée.

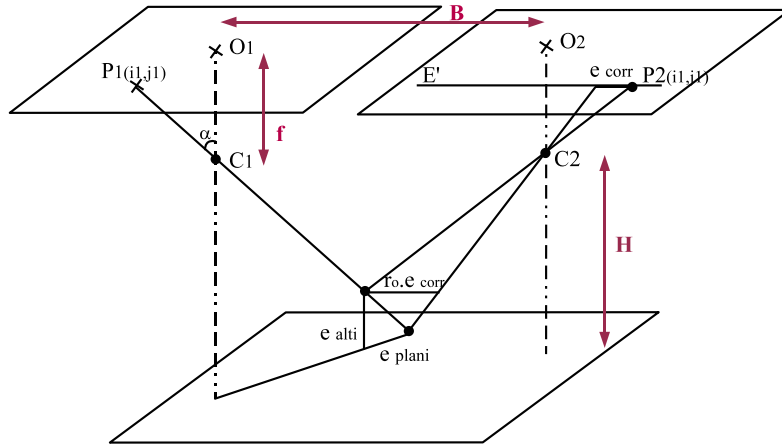


FIG. 2.4 – Erreur résultant du processus de corrélation

2.2.4 Qualité a priori d'un MNE

La qualité a priori d'un MNE se traduit par les prédictions effectuées en amont du calcul du MNE, en utilisant des informations sur le capteur, la configuration de prise de vue, ou bien encore des informations a priori de la scène.

2.2.4.1 La précision géométrique

Il est possible de définir une précision théorique du MNE à partir des paramètres d'acquisition et de la précision théorique des mesures image. Cette précision peut être approximée de la façon suivante [Vei91] :

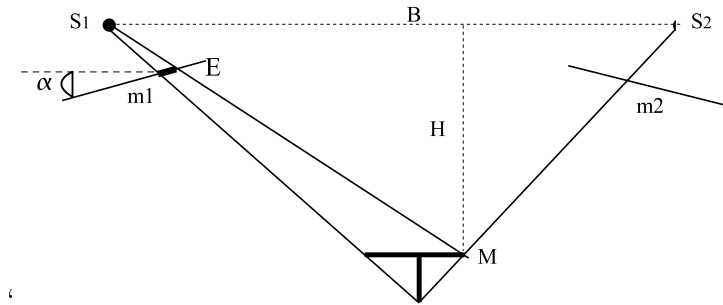


FIG. 2.5 – Précision théorique planimétrique et altimétrique

La figure 2.5 illustre deux images de sommets de prise de vue respectifs S_1 et S_2 . La base B est la distance entre ces deux sommets, H est la hauteur moyenne de vol, E l'échelle moyenne de l'image et α l'angle entre l'axe optique et la verticale.

Supposons que le point du terrain M est mesuré sur les deux images en m_1 et m_2 avec une précision σ_p , la précision planimétrique σ_{xy} correspond à la précision de mesure rapportée à l'échelle moyenne de l'image. La précision altimétrique est quant à elle inversement proportionnelle au rapport B/H (cf. equation 2.2).

$$\begin{cases} \sigma_{xy} = \sigma_p * \frac{E}{\cos(\alpha)} \\ \sigma_z = \frac{\sigma_{xy}}{B/H} \end{cases} \quad (2.2)$$

Sur les données de type simulations PLÉIADES, on obtient donc les précisions théoriques suivantes, en supposant un écart-type sur les mesures image de 0,5 pixel :

TAB. 2.1 – Précisions théoriques sur les simulations PLÉIADES

Résolution	B/H	Roulis	$\sigma_{xy}(m)$	$\sigma_z(m)$
60 cm	0.7	0°	0.30	0.43
		20°	0.32	0.46
	0.05	0°	0.30	6.00
		20°	0.32	6.39
80 cm	0.7	0°	0.40	0.57
		20°	0.43	0.61
	0.05	0°	0.40	8.00
		20°	0.43	8.51

L'approximation faite ci-dessus est valable pour un relief peu chahuté. Si l'on se place dans le contexte des MNE urbains à haute résolution, la relation approximée entre la précision planimétrique et la précision altimétrique n'est plus valable comme on le voit sur la figure suivante :

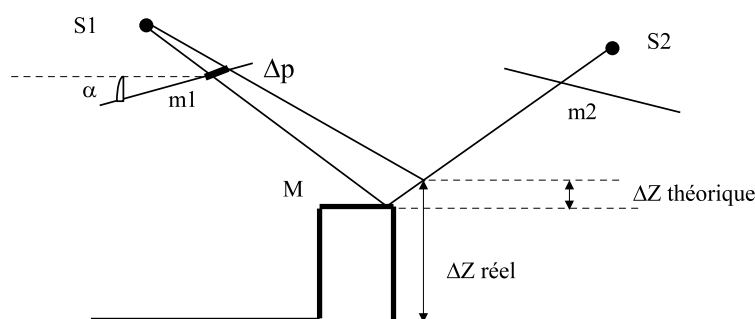


FIG. 2.6 – Précision théorique planimétrique et altimétrique

La figure 2.6 montre l'influence d'une erreur de mesure image sur la qualité de localisation altimétrique du point restitué. Une petite erreur en xy peut induire une erreur importante en z au niveau des bords de bâtiments. Cette constatation sera prise en compte, dans l'évaluation du MNE, pour la séparation de l'erreur planimétrique et de l'erreur altimétrique (cf. chapitre 8).

2.3 Les principales approches de mise en correspondance

Cette partie présente les différentes approches permettant de traiter le problème de mise en correspondance en stéréoscopie. Les différentes approches dépendent essentiellement de l'application recherchée et des contraintes utilisées. Les contraintes peuvent être liées à la stéréoscopie

et à la scène telle que la contrainte épipolaire, la contrainte d'unicité, de continuité, etc. Les contraintes peuvent être introduites par le choix de l'algorithme de reconstruction de surfaces (approche surfacique vs approche énergétique) ou bien par des contraintes images provenant de la mise en correspondance de primitives 2D (points, segments, courbes...). Les approches peuvent être séparées en trois classes distinctes : les méthodes surfaciques (**area-based**, cf. 2.3.1), les méthodes à base de primitives (**feature-based**, cf. 2.3.2) et les méthodes énergétiques (cf. 2.3.3). On a gardé la même catégorisation que C. Vestri [VD01]. L'état de l'art sera présenté dans le cadre de la stéréoscopie, certaines méthodes pouvant être appliquées dans un contexte multi-vues.

2.3.1 Les méthodes surfaciques

L'idée générale consiste à calculer une mesure de ressemblance entre les points à appairier. Les méthodes surfaciques proposent en général un point homologue pour chaque point à appairier d'où une description dense de la scène. Cependant, ces méthodes rencontrent des difficultés au niveau des discontinuités d'altitude et sont efficaces si la scène respecte certaines contraintes : la contrainte lambertienne pour que les voisinages se ressemblent, la contrainte fronto-parallèle pour minimiser les déformations inter-images. Les méthodes surfaciques sont séparées essentiellement en deux catégories : **les méthodes surfaciques locales** qui cherchent à appairier les pixels d'une image suivant des contraintes locales, et **les méthodes surfaciques globales** où l'appariement des deux images se fait en utilisant des contraintes globales sur toute la scène permettant de gérer explicitement les discontinuités et les occultations. Les approches locales et globales sont détaillées dans ce qui suit.

2.3.1.1 Les méthodes surfaciques locales : Critère de corrélation

Les méthodes surfaciques locales recherchent des couples de points homologues dans les deux images suivant les lignes épipolaires. On se fixe un point d'intérêt dans la première image et on recherche son homologue dans la seconde image en maximisant un critère de ressemblance prédéfini qui respecte les contraintes locales imposées.

- **Utilisation des critères de corrélation** : Une des méthodes classiques de mise en correspondance de points est l'utilisation d'un critère de ressemblance entre les vecteurs d'attributs correspondants aux voisinages de ces points dans l'image. Les attributs pouvant être la valeur d'intensité du pixel par exemple ou tout autre mesure discriminatoire permettant la mise en correspondance des pixels. La plupart des mesures de corrélation sont basées sur deux mesures : la distance euclidienne entre les deux vecteurs (notée SSD comme **Sum of Squared Differences**) et le produit scalaire des deux vecteurs (noté CC comme **Cross Correlation** ou Corrélation Croisée). Le principe est de calculer, pour chaque pixel et chaque valeur de disparité d possible, le score de corrélation correspondant. Les meilleurs scores de corrélation renvoient la carte de disparité finale. Les principales difficultés apparaissent avec des surfaces à forte pente à cause des distorsions entre images, avec les surfaces non texturées ou bien avec des textures répétitives car les scores de corrélation ne sont plus discriminatifs et avec les parties cachées qui ne sont pas vues dans l'une des deux images et qui ne seront pas appariées.
- **Gestion des discontinuités** : Afin de préserver les discontinuités de profondeur et les occultations, il est possible de réajuster le support de corrélation pour stabiliser l'appariement. On utilise alors des fenêtres adaptatives dont la taille est définie en fonction d'une analyse locale de la disparité courante et des intensités [Lot96]. Exploitant aussi une carte

de contours, d'autres approches utilisent une taille fixe de fenêtre munie d'un masque adaptatif défini par diffusion[Pap98], ou bien d'un masque adaptatif pondéré pour privilégier le voisinage central, qui est moins sensible aux problèmes de distorsions dus à des surfaces de fortes pentes [Cor98].

La plupart des algorithmes cités précédemment considèrent que toutes les discontinuités d'intensité correspondent à des discontinuités de profondeur, autrement dit à des discontinuités d'altitude. Ce qui n'est pas forcément le cas pour des zones fortement texturées où les discontinuités d'intensité correspondent à des changements de texture et non à un changement d'altitude. Par conséquent, la carte de disparité finale obtenue par ces méthodes n'est pas toujours dense.

2.3.1.2 Les méthodes surfaciques globales

► Approches par programmation dynamique

La programmation dynamique est une technique qui permet de résoudre les problèmes d'optimisation avec contraintes, qui consiste à minimiser une fonction de coût dans un graphe bidimensionnel (dans le cas de la stéréoscopie). Le graphe bidimensionnel constitue l'espace de mise en correspondance et on cherche à minimiser le coût global le long de la ligne des appariements dans ce graphe 2D. Ce chemin d'appariements permet de prendre en compte les contraintes d'unicité et d'ordre, tout en imposant une cohérence globale et en tenant compte des discontinuités et des occultations. La méthode est efficace pour optimiser l'appariement le long d'une ligne épipolaire ; elle est particulièrement répandue pour les points de contour ou les intervalles [OK85, Bai97]. Le coût local défini pour chaque paire de pixels doit être discriminant pour renvoyer un faible coût pour les bons appariements. On peut définir par exemple un coût inverse au score de corrélation. H.Ishikawa et D.Geiger utilisent dans [IG98a] une fonction de coût qui prend en compte la plupart des contraintes existantes et qui modélise aussi les discontinuités et les occultations.

D'autres travaux tiennent compte du voisinage extérieur des lignes épipolaires. Y.Ohta et T.Kanade proposent dans [OK85] de combiner deux processus de recherche utilisant la programmation dynamique ; une recherche inter-lignes épipolaires et une recherche intra-lignes épipolaires.

S. Intille et A.Bobick proposent dans [IB95] une méthode de mise en correspondance basée sur la programmation dynamique dans les images des espaces de disparités DSI (Disparity Space Image).

► Autres approches par surfaces

Les méthodes exposées ci-dessus permettent d'avoir un ensemble de points 3D non structuré. Leur interprétation directe est difficile et il est souvent nécessaire de les structurer en données plus compactes afin de mieux interpréter la scène. Dans la littérature, on peut distinguer deux stratégies de reconstruction de surfaces à partir d'un ensemble de points 3D. Les premières approches consistent à regrouper des primitives 3D (points 3D, segments 3D...) appartenant à une même surface puis à estimer cette surface. Cette stratégie est basée sur les primitives et sera détaillée dans la section 2.3.2. La deuxième stratégie consiste à quadriller l'espace de manière régulière, estimer les surfaces locales puis les regrouper [Fua97, VD01, TGD03].

2.3.2 Les méthodes à base de primitives

En milieu urbain, la présence de superstructures (cheminées, chiens assis) ou de motifs répétitifs rendent les techniques de corrélation pixel à pixel sensibles. L'extraction de primitives 3D fiables devient alors un élément clef et souvent une étape préliminaire pour permettre une description sémantique de la scène et faciliter la génération d'hypothèses de bâtiments. L'utilisation des primitives comporte deux principaux avantages : la réduction du volume de données à traiter et l'apport d'informations supplémentaires liées à la particularité des primitives extraites. Ces primitives 3D sont en général des coins [FKL⁺98], des plans ou des segments 3D [FL99, WA96].

2.3.2.1 Choix des primitives

Les propriétés nécessaires à l'appariement des primitives sont l'invariance par rapport au point de vue, le pouvoir discriminant, la stabilité, la précision de la localisation et la densité dans l'image. Ces propriétés sont rarement toutes vérifiées. Il faut choisir un bon compromis en fonction de l'objectif fixé. Les primitives les plus courantes sont les points d'intérêt, les chaînes de contour, les segments linéaires et les régions homogènes. Il est également possible d'apparier des jonctions [HK86] ou des rectangles [MN89]. Naturellement, plus les primitives sont complexes, plus l'appariement est fiable, mais moins la détection est exhaustive. Deux difficultés seront rencontrées avec l'utilisation des primitives 3D, la sous-détection des primitives qui ne seront pas en nombre suffisant pour reconstruire une scène dense et la sur-détection c'est-à-dire des primitives qui ne correspondent à aucune information 3D et qui risquent de fausser la reconstruction finale.

2.3.3 Les méthodes énergétiques

Les approches énergétiques constituent une troisième classe de méthodes permettant de résoudre le problème de mise en correspondance comme un problème de minimisation d'une énergie globale. Il existe trois principales techniques pour résoudre le problème :

- **L'approche bayésienne**, qui cherche à maximiser la probabilité d'appariement. Le problème est formulé de façon probabiliste en modélisant le processus de formation des images et la scène 3D. Les algorithmes bayésiens se distinguent en général par le choix des modèles utilisés et le choix de la méthode d'estimation. Soit S , la géométrie d'une scène 3D, un couple stéréoscopique I_d et I_g sur la scène. Le problème consiste à rechercher la géométrie la plus probable de la scène, connaissant les deux images. On définit $P(S|I_d, I_g)$, la **probabilité a posteriori** de l'état de la scène, donné par les mesures comme suit :

$$P(S|I_d, I_g) = \frac{P(I_d, I_g|S)P(S)}{P(I_d, I_g)} \quad (2.3)$$

- $P(S|I_d, I_g)$ est la probabilité de la scène 3D, S , reconstruite à partir des images I_d et I_g .
- $P(I_d, I_g|S)$ est le modèle de formation des images connaissant la scène, appelé aussi **modèle des données**.
- $P(S)$ est le **modèle a priori** de la scène.
- $P(I_d, I_g)$ ne dépend pas de la géométrie de la scène. La probabilité qui lui est associée est donc considérée comme constante ou indépendante.

L'énergie fonctionnelle du modèle des données représente le processus de formation des images. Pour que ce processus soit correctement modélisé, il faut que l'énergie prenne en

compte les occultations entre images. Dans [Bel96], l'auteur propose un modèle de formation des images dérivé de la fonction de disparité du modèle cyclopéen qui prend en compte les occultations. Dans [GLY95], les auteurs modélisent le processus de mise en correspondance des images au lieu du processus de formation des images, ce qui leur permet de gérer les occultations.

L'énergie fonctionnelle du modèle a priori représente les connaissances dont on dispose sur la scène 3D. Dans [Bel96], l'auteur traite trois modèles de scène allant du plus simple au plus complexe. La première représentation considère que la scène est constituée uniquement de surfaces lisses. Le deuxième modèle prend en compte des discontinuités de profondeur dans la scène 3D. Le dernier modèle proposé, prend en compte des discontinuités d'orientation en supposant que certains objets peuvent être linéaires par morceaux le long des lignes épipolaires.

Les auteurs réalisent par la suite une minimisation de l'énergie composée du modèle a priori et du modèle de données. Dans [Bel96, GLY95], les auteurs utilisent la programmation dynamique. Dans [RC98, IG98a], ils utilisent une minimisation d'énergie à base de flots de graphe.

- **L'approche variationnelle**, qui transforme le problème de mise en correspondance en un problème de minimisation d'énergie [RD95, FK98, GF99]. Les auteurs introduisent les équations à dérivées partielles *EDP*. Dans [FK98], les auteurs définissent un ensemble d'*EDP* pour faire évoluer une surface 3D initiale vers les objets de la scène. La résolution du système est réalisée à l'aide de surfaces de niveaux (**Level Sets**) permettant de prendre en compte les changements de topologie. Les auteurs proposent une implémentation 2D de la méthode. Dans [GF99], les auteurs proposent une nouvelle *EDP* et une implémentation tridimensionnelle.
- **L'approche à base de flots de graphe**, qui consiste à chercher les coupes de capacité minimale dans des graphes [RC98, IG98a, BVZ99, Vek99, KZ01] afin de minimiser une énergie globale traduisant un certain nombre de contraintes (régularité de surface, discontinuités ou occultations). Cette dernière approche sera détaillée dans le chapitre 7.2.

2.4 Les modèles de bâtiments

En stéréoscopie simple, avec un processus automatique d'extraction de bâtiments, on est rapidement confronté aux problèmes des occultations qui rendent la reconstruction 3D de bâtiments très complexe en milieu urbain dense. La plupart des travaux portant sur la stéréoscopie, utilisent des modèles simples de bâtiments : segmentation en facettes planes sans topologie [CJC01]. D'autres auteurs ont proposé des modèles plus complexes : Dans [PCJC98], les algorithmes fournissent des modèles prismatiques. Dans [FL99], l'auteur propose des modèles structurels basés sur un formalisme sous forme d'isomorphisme de graphes (graphe de données et graphe modèle) qui permet de pallier aux problèmes de sous-détection de primitives.

Dans la plupart des cas, les auteurs supposent que les bâtiments sont polyédriques c'est-à-dire qu'ils peuvent être représentés par des facettes planes. Cette représentation permet de décrire la plupart des bâtiments. Dans ce qui suit, on fera le tour des différents modèles de bâtiments présents dans la littérature. Les modèles diffèrent énormément en fonction des données dont on dispose, c'est-à-dire en fonction du nombre d'images, de la résolution, etc. Dans cette énumération, on ne se restreindra pas au cas de la stéréoscopie simple.

2.4.1 Caricature de surface

Comme première classe, on retrouve la caricature de surface où les algorithmes ne fournissent pas une représentation vectorielle du bâtiment à proprement dite. Dans [Ves00, CJC01], les auteurs réalisent une classification sol/ sursol suivie d'une étape de focalisation sur l'objet à reconstruire et réalisent finalement une segmentation du MNE en facettes planes. D'autres approches sont basées sur les contours actifs [RD95], où l'on modélise des surfaces 3D à l'aide d'équations à dérivées partielles et que l'on fait évoluer pour "coller" au mieux aux objets de la scène. On notera que dans ces approches, aucune topologie n'est reconstruite.

2.4.2 Approches par modèles paramétriques

Les modèles paramétriques décrivent la forme des objets avec un jeu minimum de variables (cf. figure 2.7(a)). Pour ces modèles, la topologie est fixée, par contre la géométrie, la localisation et l'orientation de l'objet sont variables. Ils décrivent des formes simples : bâtiments rectangulaires à toits plats [VC02b], bâtiments non orthogonaux à toits plats, bâtiments rectangulaires avec des toits en pente [Wei97]. Ces travaux se limitent également à des bases rectangulaires [ODZ03], travaux de Vinson et Cohen [VC02a, VC02b, SF01]. L'avantage de ces modèles est qu'ils permettent d'avoir une modélisation complète du bâtiment grâce au volume paramétrique. Par contre, on ne pourra pas modéliser des formes complexes de bâtiments et les bâtiments reconstruits sont très contraints par le modèle initial. La stratégie adoptée en pratique, consiste à choisir une position et une orientation initiale, projeter le modèle paramétrique dans les images et à ré-estimer les paramètres du modèle jusqu'à ce que l'enveloppe projetée du bâtiment corresponde aux contours de l'image.

2.4.3 Approches génériques

Les modèles génériques permettent de décrire toutes les formes de bâtiments. Ils résultent de l'assemblage de différentes primitives. On distingue trois types de modèles génériques présentés du plus contraint au moins contraint :

- **Les modèles prismatiques** : (cf. figure 2.7(b)) supposent que le bâtiment peut être décrit par un polygone dans le plan du sol, défini à une altitude fixe. Le modèle est généralement déduit d'un assemblage de segments. Le toit est horizontal et les façades sont verticales. Le nombre de paramètres du modèle est défini par rapport au nombre de sommets du polygone [CJC⁺98].
- **Les modèles structurels** : Le modèle de bâtiment est décomposé en primitives simples qui sont associées par union, soustraction, intersection. Ceci permet de restreindre la bibliothèque de formes à reconstruire [FL99, FKL⁺98].
- **Les modèles polyédriques** décrivent les bâtiments par assemblage de surfaces planaires. Dans [NH97, JPDPM00], le modèle final est obtenu par fusion et intersection de plans 3D. Dans [NN01], les auteurs proposent une technique de groupement perceptuel qui permet de générer des hypothèses de toits de bâtiments. Dans [BZ99], on recherche des demi-plans 3D pivotant autour des segments 3D extraits.
La description du modèle est un peu plus complexe mais permet de décrire la quasi-totalité des bâtiments. On peut aussi introduire des contraintes telles que : symétrie des toits, ver-

ticalité des façades, etc.

L'inconvénient de ces approches génériques est principalement leur manque de robustesse face aux problèmes de sous-détection de primitives. On retrouve dans [Ves00], un état de l'art détaillé et très complet concernant les approches de reconstruction de bâtiments par équipe de recherche.

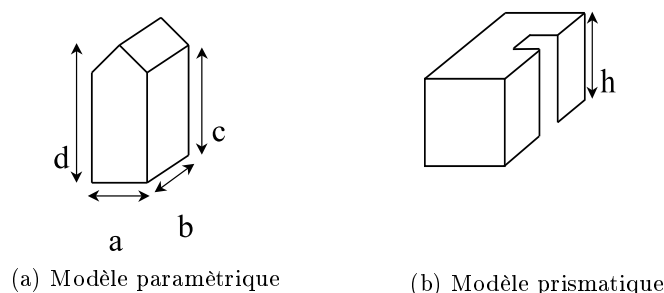


FIG. 2.7 – Modèles de bâtiments

2.4.4 Bilan

Pour extraire de l'information à partir des données initiales, deux stratégies sont généralement déployées : soit une approche descendante à base de modèles, soit une approche ascendante à base de primitives extraites des images. Dans l'approche descendante, une bibliothèque de modèles est définie et on essaye d'apparier au mieux ces modèles à l'ensemble des primitives extraites. Ces approches (**top-down**) sont limitées par la librairie de modèles qui n'est pas assez générique et sont souvent très combinatoires, mais elles ont l'avantage d'être plus robustes aux sous-détections de primitives. Dans l'approche ascendante (**Bottom-Up**), un ensemble de primitives fiables est extrait des images et le modèle de bâtiments est construit sans aucune connaissance a priori. Les problèmes de sous-détections sont résolus par groupement perceptuel des primitives et par des contraintes locales. Ces approches sont plus génériques.

La fusion des deux stratégies fournit une stratégie mixte qui consiste à émettre plusieurs hypothèses et en valider une (**Hypothesis-and-verify**), les hypothèses sont fournies par une stratégie ascendante et validées par une stratégie descendante en imposant des modèles de contraintes locales. Ces approches mixtes ont l'avantage de ne pas restreindre la forme du bâtiment a priori, mais donnent la possibilité d'appliquer un modèle de contraintes globales.

Dans ce chapitre, on a présenté les différents niveaux d'interprétation ou de modélisation de scènes urbaines à partir d'un couple stéréoscopique, indépendamment de la résolution de l'image. En fonction des données en entrée, un niveau d'interprétation spécifique pourra être choisi ; des modèles de bâtiments très détaillés, des modèles simplifiés ou bien des éléments de surfaces pourront être reconstruits.

Chapitre 3

Stratégie globale

3.1 Introduction

Les modèles de bâtiments ainsi que les stratégies de reconstruction dépendent du contexte de travail et des données dont on dispose. On se place dans le cadre de cette thèse dans un contexte satellitaire très haute résolution. Le but étant d'évaluer la modélisation 3D de scènes urbaines à partir des images satellitaires submétriques (à très haute résolution). On se place en stéréoscopie simple. Une stratégie mixte a été adoptée. Elle comporte deux parties principales : L'extraction de primitives 3D et l'utilisation de ces primitives comme contraintes dans un processus de modélisation globale de la scène.

Ce chapitre permettra d'introduire et de justifier la stratégie adoptée.

3.2 Objectif : Modélisation de surfaces 3D

À cause de la faible résolution et de la qualité des images, les détecteurs de primitives souffrent des problèmes de sous-détections et de surdétections. La reconstruction d'un modèle polyédrique à partir des primitives extraites devient alors très compliquée. De plus, la fiabilité des modèles ainsi reconstruits reste faible de par le manque d'images disponibles pour la vérification et la validation des hypothèses de primitives.

Par conséquent, on propose une modélisation **dense** de surfaces 3D "**hybrides**" (raster/vecteur) (cf. figure 3.1) qui combine les primitives 3D extraites et une information raster aux endroits où l'on manque d'information vectorielle fiable. Ce produit hybride (cf. figure 3.1) propose ainsi une modélisation dense de la scène avec différents niveaux de description des bâtiments suivant la fiabilité des primitives 3D extraites. Le but sera de combiner les informations extraites qui se trouvent nécessairement à plusieurs niveaux de description (points 3D, segments 3D et plans 3D) afin d'obtenir une description de la scène permettant d'exploiter à chaque endroit, les primitives du niveau le plus élevé que l'on a pu reconstruire de manière fiable.

3.3 Démarche

Le problème de reconstruction de bâti à partir d'images aériennes a fait l'objet de nombreuses recherches (cf. section 2.4). On peut distinguer deux types d'approches. Les approches à partir de modèles utilisent une base prédéfinie de modèles de bâtiments (définis soit sous forme paramétrique, soit sous forme structurelle c'est à dire par agrégation de formes simples) et recherchent

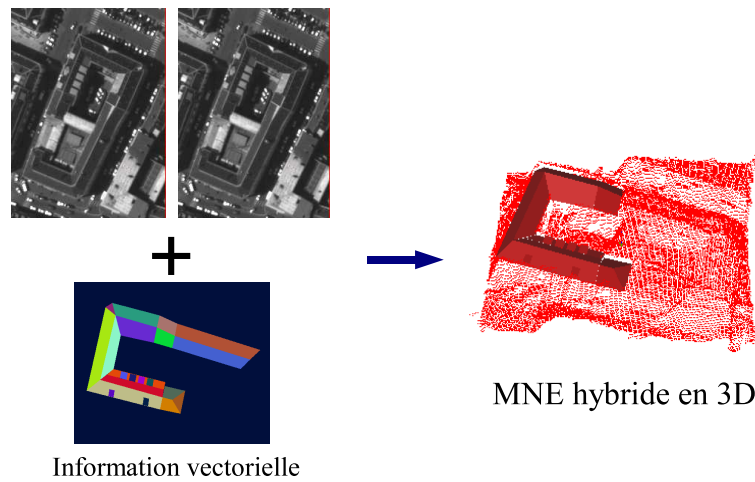


FIG. 3.1 – MNE hybride raster/vecteur

à partir des données le meilleur modèle correspondant dans la base. Ces approches sont très robustes aux détecteurs de primitives mais sont, de façon évidente, peu génériques dans la mesure où un modèle ne peut pas être reconstruit s'il n'est pas dans la base. Les approches à partir des primitives fonctionnent par détection de primitives puis leur agrégation pour fournir la reconstruction finale. Cette modélisation est très générale sous la forme d'un polyèdre quelconque mais ces approches sont beaucoup moins robustes aux échecs inévitables des détecteurs de primitives.

Une stratégie mixte a donc été choisie (cf. figure 3.2). Un ensemble d'hypothèses concurrentes est émis grâce aux primitives pour le modèle final, gardant ainsi la généralité de ce type d'approches et une approche descendante de modélisation globale de la scène 3D, intégrant des contraintes locales pour rester robuste aux erreurs des détecteurs de primitives.

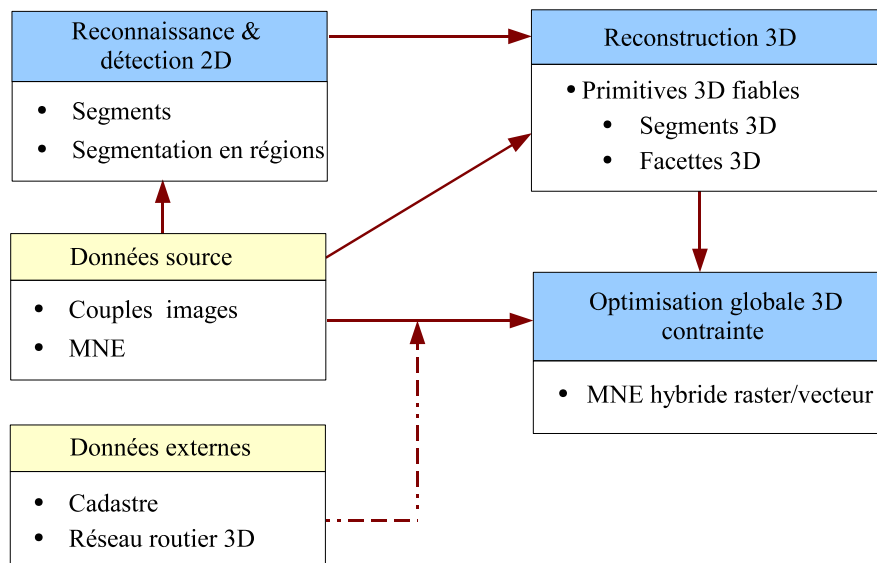


FIG. 3.2 – Stratégie globale

Deux représentations de la surface 3D finale seront proposées : une représentation raster, c'est-à-dire une grille régulière où la valeur de chaque point renvoie son altitude terrain, ce qui correspond à un Modèle Numérique de Surface régularisé avec la carte de labels des primitives retenues dans la coupe. La deuxième représentation sera une représentation hybride en 3D (cf. figure 3.1) où les primitives 3D retenues (segments 3D, facettes 3D) seront représentées en vectoriel. Aux endroits où l'on manque d'information vectorielle fiable, l'information raster sera utilisée et représentée sous forme de grille 3D.

Ce MNE hybride pourra être utilisé dans de nombreuses applications telles que la génération d'orthoimages ou la réalité virtuelle.

On détaillera dans les sections suivantes les deux grandes parties de cette stratégie qui sont essentiellement la reconstruction des primitives et le processus de modélisation globale.

3.3.1 Reconstruction des primitives

Notre objectif consiste à modéliser principalement les toits de bâtiments en milieu urbain dense en utilisant deux types de primitives 3D : les segments 3D et les facettes 3D. On ne cherchera pas à détecter les façades qui sont très difficiles à extraire en milieu urbain dense et d'autant plus en stéréoscopie simple, en visée quasi-verticale. Ces primitives permettront de contraindre le problème d'optimisation globale pour la modélisation de la surface 3D finale. Par conséquent, les primitives extraites devront être **fiables**. On privilégiera la fiabilité par rapport à l'exhaustivité de la reconstruction.

Les deux types de primitives, linéique et surfacique, présentent l'avantage d'être complémentaires : les segments permettent de bien structurer la scène. Ils correspondent surtout aux discontinuités et présentent l'avantage d'avoir une bonne localisation planimétrique. Les facettes 3D sont, quant à elles, plus fiables pour leur précision altimétrique, mais par contre moins nombreuses. En effet, les primitives 2D (régions) étant plus complexes dans ce dernier cas, leur appariement est moins exhaustif. De plus, la délimitation des facettes 3D n'est pas très précise en général, cet inconvénient sera compensé par la bonne localisation des segments 3D.

Les hypothèses de primitives sont émises par une stratégie ascendante à partir des images ou bien d'un MNE généré à partir des images. Les processus d'extraction des primitives devront tirer profit du faible rapport B/H du couple stéréoscopique. Ce paramètre a influencé le choix des méthodes et algorithmes décrits dans ce manuscrit. L'algorithme de détection des segments 3D, utilisant une contrainte photométrique, sera présenté dans le chapitre 4. Pour les surfaces planes, deux méthodes d'extraction de facettes 3D seront présentées. La première est basée sur les segments 3D extraits. La délimitation des facettes sera réalisée par relaxation probabiliste en utilisant les scores de corrélation et en mettant ainsi à profit le faible rapport B/H (cf. chapitre 5). La deuxième méthode est basée sur les régions et a l'avantage de fournir directement une description tridimensionnelle de la scène par des facettes 3D. La mise en correspondance des régions est facilitée grâce aux faibles distorsions géométriques entre les deux images (cf. chapitre 6).

La complémentarité des primitives extraites permet de mieux contraindre le problème d'optimisation (cf. section 3.3.1). On privilégiera la fiabilité des primitives extraites par rapport à l'exhaustivité, le problème de sous-détection sera résolu par le processus d'optimisation globale. On proposera également différents processus d'extraction de primitives afin de fournir plusieurs hypothèses concurrentes.

3.3.2 Optimisation globale

Les détecteurs de primitives souffrant de problèmes de sous-détections et surdétections, il est nécessaire de réaliser une modélisation globale de la scène. Le problème de modélisation de surfaces 3D peut être formulé comme un problème de minimisation d'énergie par flots de graphes (cf. chapitre 7.3), qui serait contraint par les primitives extraites. Le graphe 3D hybride sera construit à partir d'un volume de corrélation sur toute la scène et des primitives 3D extraites. La surface finale sera obtenue par recherche de la coupe de capacité minimale dans le graphe 3D hybride.

Cette étape a pour but d'exploiter au mieux l'information vectorielle extraite ; les facettes 3D permettent de régulariser les toits de bâtiments. Les segments 3D modélisent les discontinuités. On peut également envisager la modélisation du sol pour contraindre l'optimisation globale. Le sol pourra être obtenu par classification sol/sur-sol à partir des données sources telles que le MNE ([Bai97]) ou bien à partir de données externes telles que le réseau routier, les limites de gouttières des bâtiments disponibles sur la BD TOPO[©] ou encore les limites cadastrales.

Ce processus d'optimisation globale permet également de pallier au problème de sous-détection des primitives grâce au volume de corrélation initial qui fournit une information **dense** sur la scène d'une part, et d'autre part de résoudre le problème de surdétection des primitives en validant une seule des hypothèses de facettes concurrentes lors de la coupe.

Le système devra être **indépendant** des processus d'extraction des primitives 3D, **capable de gérer les hypothèses concurrentes** de surfaces planes provenant de différents détecteurs et **ouvert** pour intégrer facilement toute donnée externe qui permettrait de contraindre le problème de l'optimisation globale.

3.3.3 Focalisation

La délimitation de zones de focalisation permettrait de limiter la combinatoire du problème posé et de ne traiter que les bâtiments dans les scènes urbaines. Le problème de focalisation sur les bâtiments n'a pas été traité dans le cadre de cette thèse. Des petites zones de travail sont choisies manuellement par l'utilisateur. On peut aussi exploiter des résultats de classification sol/sur-sol pour focaliser sur les bâtiments [CJC01, BSZF99].

La focalisation peut également être réalisée par des données externes telles que le réseau routier ou les limites cadastrales. Le réseau 3D extrait de la BD TOPO[©] permet de délimiter certains îlots de bâtiments. Ce réseau n'est pas toujours fiable et il est très difficile de généraliser son utilisation à toutes les scènes. Sur un plan pratique, l'introduction des limites cadastrales dans le processus serait le meilleur moyen de focalisation précise sur les bâtiments.

Les techniques de mise en correspondance, de reconstruction de primitives 3D ainsi que la modélisation de la scène 3D raster/vecteur ne nécessitent pas des zones de focalisation. Toutefois, la prise en compte de telles zones pourrait facilement être envisagée et limiterait la combinatoire des problèmes traités.

Dans la suite du manuscrit, les deux parties principales de la stratégie seront développées ; la reconstruction de primitives 3D et le processus d'optimisation globale.

Troisième partie

Reconstruction de primitives 3D

Chapitre 4

Reconstruction de segments 3D

Dans ce chapitre, l'algorithme d'extraction de segments 3D sera détaillé. Des simulations ont été réalisées en fonction de l'orientation des segments, du bruit dans les images, du rapport B/H afin de quantifier la précision théorique. Les résultats sur les images réelles seront également présentés.

4.1 Introduction

L'apport des segments 3D est essentiel pour la modélisation de scènes urbaines [NN01, BZ99, WA96]. Les segments correspondent en général aux bords des objets physiques. Ils marquent les discontinuités, les bords de bâtiments, les faites de toits, etc. Leur intérêt réside dans leur capacité à structurer la scène. De plus, la faible densité de l'information à traiter permet de limiter la combinatoire et accélère les algorithmes.

La mise en correspondance des segments est un problème difficile dû principalement aux défauts d'extraction des segments dans les images. En effet, suite aux changements de conditions de prise de vue, un segment peut être sursegmenté ou bien caché dans la seconde image.

Avec un faible rapport B/H du couple stéréoscopique, les distorsions géométriques entre les images sont réduites, ce qui permet de discriminer efficacement les appariements potentiels. Cependant, l'un des inconvénients de la stéréoscopie simple est le fait que la reconstruction de segments 3D le long des lignes épipolaires n'est pas fiable. Un compromis sera à trouver entre l'exhaustivité de la détection et la fiabilité de la reconstruction.

Dans notre stratégie, les segments permettront, dans un premier temps, d'extraire des plans 3D qui leur sont rattachés (cf. chapitre 6). Pour cette étape, seuls les segments 3D fiables en altimétrie seront retenus. Dans un deuxième temps, les segments 3D seront exploités principalement pour la fiabilité de leurs coordonnées planimétriques pour contraindre le problème d'optimisation globale (cf. chapitre 7.3). Pour cette application, on privilégiera l'exhaustivité de la reconstruction à la précision altimétrique.

4.2 État de l'art

La détection et la reconstruction de segments 3D ont fait l'objet de nombreux travaux depuis des années. Plusieurs approches ont été proposées pour apparier les segments en stéréoscopie. La difficulté dans le cas de deux images est que la contrainte géométrique est réduite à une contrainte de recouvrement. Pour pallier à ce problème, deux types de stratégies ont été développés dans la littérature : la première repose sur la mise en correspondance de groupes de segments, ce qui

introduit des contraintes fortes. Cette idée de mise en correspondance par des graphes se retrouve chez de nombreux auteurs [Aya89, HS89]. Ces méthodes sont a priori plus fiables, mais sont souvent sensibles aux erreurs de détection des segments qui modifient les apparences des graphes d'une image à une autre, et ont une grande complexité. La deuxième stratégie repose sur l'introduction d'attributs géométriques ou radiométriques divers tels que l'orientation, la longueur, etc. [ZF92, GBB98] ou le voisinage radiométrique du segment [SZ97]. L'ensemble de ces attributs est souvent sensible aux erreurs d'extraction de segments dans les images, mais également aux conditions de prises de vue.

Dans notre cas, compte tenu du faible B/H , les distorsions entre les images sont réduites ainsi que les parties cachées. Les scores de corrélation entre les deux images sont alors fiables. On retiendra la deuxième stratégie d'appariement des segments qui consiste à combiner des contraintes géométriques qui sont essentiellement le recouvrement des segments 3D et leurs orientations par rapport à la ligne de base des centres de prise de vue (cf. section 4.3.3.1) et des contraintes photométriques (voisinage radiométrique des segments) pour tirer profit du faible rapport B/H du couple stéréoscopique.

4.3 Algorithme

Cette section détaille l'algorithme de reconstruction des segments 3D (cf. figure 4.1). L'approche est basée sur l'appariement des segments 2D dans l'espace image. L'étape de mise en correspondance est réalisée de manière classique en utilisant les contraintes géométrique et photométrique. L'algorithme fournit un ensemble d'hypothèses d'appariement de segments 2D. Cet ensemble d'hypothèses sera finalement élagué suivant la contrainte d'unicité. Les deux images sont traitées de manière symétrique. Finalement, une étape de validation des segments 3D obtenus (appartenance à une bande de tolérance, orientations) permettra d'obtenir la reconstruction finale.

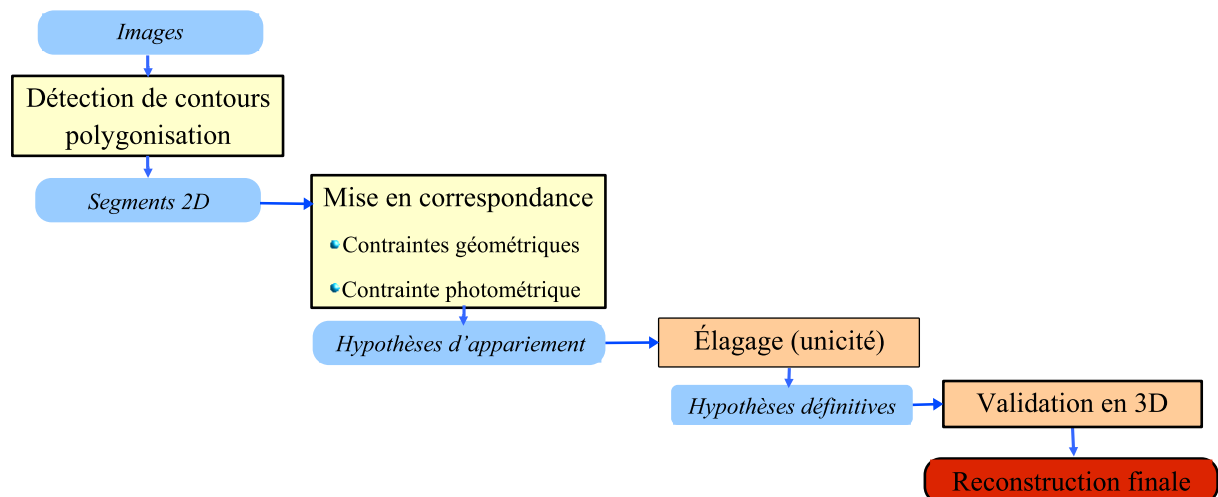


FIG. 4.1 – Algorithme de reconstruction de segments 3D

4.3.1 Extraction des segments dans les images

L'extraction des segments 2D dans les images est réalisée en utilisant l'algorithme décrit dans [Tai04]. Dans un premier temps, une extraction de contours est réalisée à l'aide d'un détecteur de Canny Deriche [Der87] sur les images suréchantillonnées (ce qui permet d'avoir une localisation subpixelaire des contours), suivi d'un seuillage hystérésis à deux paramètres S_b et S_h . Les contours sont chaînés et polygonalisés. On utilise pour cela, un processus de fusion itératif basé sur le résidu maximum de la régression orthogonale. Dans ce qui suit, la mise en correspondance des segments 2D et la reconstruction finale des segments 3D seront détaillées.

4.3.2 Mise en correspondance

Afin de réduire la combinatoire, on applique sur les extrémités des segments les contraintes épipolaire et altimétrique. Le MNE (Modèle Numérique d'Élévation), obtenu par des techniques de corrélation [Bai97], fournit l'altitude minimale et maximale de chaque pixel. Ce qui permet de réduire l'espace de recherche (cf. figure 4.2).

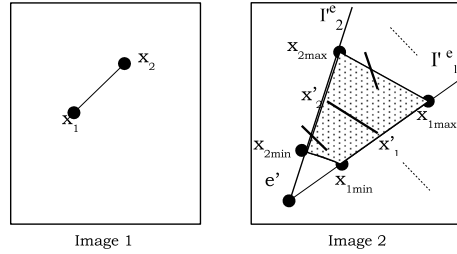


FIG. 4.2 – Contrainte épipolaire et altimétrique

l_1^e et l_2^e constituent les lignes épipolaires correspondant respectivement à x_1 et x_2 dans la seconde image. e' constitue l'épipoles et se trouve à l'intersection de toutes les lignes épipolaires. L'homologue du segment $[x_1x_2]$ se situe donc entre les deux lignes épipolaires correspondantes à ses extrémités. La contrainte altimétrique permet ensuite de réduire l'espace de recherche à un polygone.

4.3.2.1 Contrainte géométrique

En stéréoscopie, la contrainte géométrique est réduite à une contrainte de recouvrement. Le score géométrique sera alors égal au taux de recouvrement des segments appariés, sur la droite 3D correspondante. Les associations de segments dont le taux de recouvrement est $< 50\%$ sont rejetées.

Soit C_1, C_2 les deux centres de prise de vues. S_1, S_2 deux segments homologues, la droite 3D correspondante est calculée comme l'intersection des deux plans perspectifs $\Pi(C_1, S_1)$ et $\Pi(C_2, S_2)$ (cf. figure 4.3).

4.3.2.2 Contrainte photométrique

Pour chaque paire de segments satisfaisant la contrainte géométrique ci-dessus, on applique une contrainte photométrique qui prend en compte le voisinage du segment. La géométrie épipolaire permet d'obtenir une correspondance pixel à pixel entre les segments homologues. Le score

photométrique est obtenu en moyennant les scores de corrélation individuels de tous les pixels du segment avec leurs homologues respectifs.

$$score_{Ph}(S, S') = \frac{1}{n} \sum_{i=1}^n Corr(P_i, P'_i) \quad (4.1)$$

avec

- n le nombre de pixels du segment S , pris comme segment de référence.
- P_i les pixels du segment S , P'_i étant les points correspondants par la contrainte épipolaire.
- $Corr(x, x')$ est la corrélation centrée normalisée entre deux points homologues x et x' .

4.3.2.3 Associations finales

Le score d'appariement final prend en compte les deux contraintes géométrique et photométrique comme suit :

$$\begin{cases} Score_{Final} = \lambda_G * Score_G + \lambda_{Ph} * Score_{Ph} \\ avec \lambda_G + \lambda_{Ph} = 1 \end{cases} \quad (4.2)$$

En pratique, on donne plus de poids au score photométrique puisque les scores de corrélation sont fiables grâce au faible rapport B/H du couple. Le choix des valeurs de pondération n'est pas critique, les bons appariements renvoyant un bon score pour les deux contraintes. Les deux images sont traitées de façon symétrique et un ensemble d'associations 1 – 2 (image 1 considérée comme image de référence) et réciproquement 2 – 1 est fourni. L'ensemble d'associations consistant est retenu suivant le schéma de "the winner takes all" pour garder un seul homologue pour chaque segment.

4.3.3 Reconstruction des segments 3D et validation

On considère un couple de segments homologues, on projette les extrémités 2D sur la droite 3D correspondante et on applique une stratégie d'union qui consiste à garder les extrémités les plus éloignées. On obtient ainsi le segment 3D final (cf. figure 4.3). Cette stratégie d'union entraîne la nécessité de valider les extrémités du segment reconstruit a posteriori. On utilise pour cela, un volume de recherche dérivé du MNE par dilatation, (cf. figure 4.3) auquel doivent appartenir les extrémités des segments.

Le MNE dont on dispose en entrée n'est pas précis, mais robuste. Ces altitudes minimales et maximales sont par conséquent fiables. Elles sont utilisées pendant l'étape de mise en correspondance pour restreindre l'espace de recherche pour la contrainte épipolaire et fournissent une bande de tolérance autour du MNE pour valider la reconstruction 3D des segments.

4.3.3.1 Contrainte d'orientation

Les segments reconstruits sont ensuite élagués suivant une contrainte d'orientation par rapport à la base des centres de prise de vue. Comme on le verra dans la section 4.4, des simulations ont montré l'influence de l'orientation du segment sur la précision de sa reconstruction en 3D. La figure 4.5 montre que la précision de reconstruction, même au niveau planimétrique, est rapidement dégradée quand le segment est parallèle à la base de prise de vue. À partir de cette courbe théorique, un seuil de 15° est appliqué en pratique. L'angle des segments 3D par rapport à la base doit être supérieur à ce seuil pour assurer leur fiabilité.

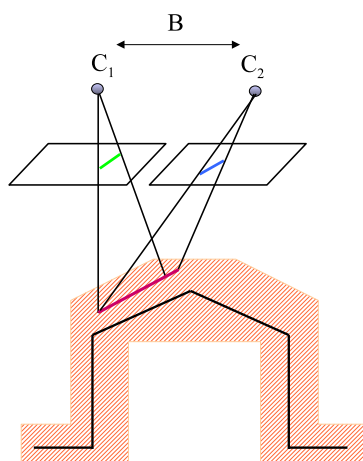


FIG. 4.3 – Reconstruction et validation de segments 3D

4.4 Résultats des simulations

Dans un premier temps, des simulations permettent de mesurer les précisions théoriques des segments 3D reconstruits et l'impact de différents paramètres tels que l'orientation des segments, le bruit dans les images ou encore le rapport B/H du couple, sur la qualité de la reconstruction 3D des segments.

4.4.1 Simulation des segments 3D

La simulation de segments 3D suivant différentes configurations permet d'évaluer la précision de la reconstruction 3D. Un segment 3D horizontal de longueur $50m$ est simulé. Le segment est projeté dans les deux images. Les segments 2D sont échantillonnés à la résolution de l'image. Les points constitutifs des segments 2D sont bruités avec un bruit gaussien isotrope d'écart type σ . Le segment 3D est ensuite reconstruit à partir des segments bruités. On réalise 10000 tirages aléatoires pour chaque valeur de bruit. La précision de reconstruction du segment simulé correspond à la différence entre les vecteurs normalisés du segment reconstruit et du segment simulé. Elle est représentée en mètres. Les précisions altimétrique et planimétrique sont calculées séparément.

4.4.1.1 Précision en fonction de l'orientation :

Les premières simulations ont été réalisées en fonction de l'angle du segment 3D simulé par rapport à la base entre les deux centres de prise de vue. Le rapport B/H du couple est égal à 0.2.

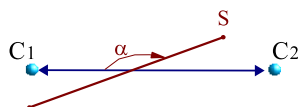


FIG. 4.4 – Segment 3D simulé

La figure 4.5 illustre la moyenne de précision obtenue sur les 10000 tirages ainsi que l'écart type des mesures.

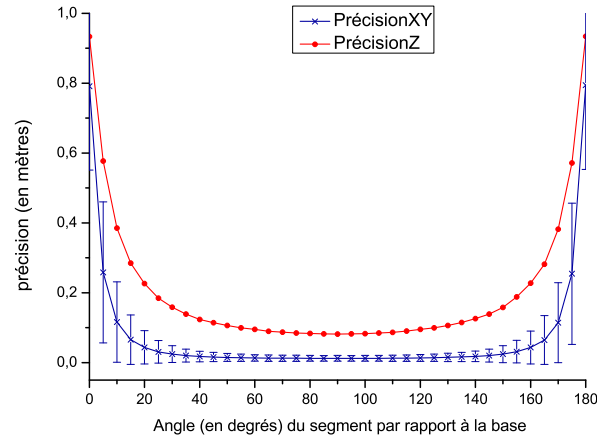


FIG. 4.5 – Précision de reconstruction suivant l'angle du segment par rapport à la base des centres de vue

On retrouve bien le résultat attendu, la meilleure précision de reconstruction est obtenue pour un segment horizontal, perpendiculaire à la base des centres de prise de vue. Dans un intervalle de 15° par rapport à la base, la précision des segments n'est pas fiable. En pratique, l'angle du segment par rapport à la base sera également un critère de sélection des segments 3D finaux.

On se place pour le reste des simulations dans la meilleure configuration pour l'angle par rapport à la base des centres de prise de vue c'est-à-dire à 90° degrés.

4.4.1.2 Précision en fonction du bruit :

La figure 4.6 montre la précision de la reconstruction en fonction de l'écart type du bruit sur les segments dans les images.

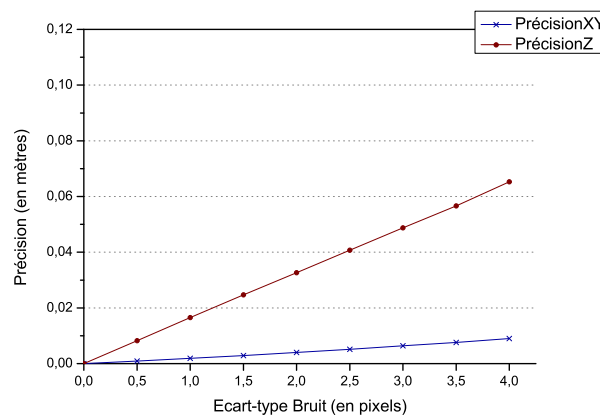


FIG. 4.6 – Précision de reconstruction en fonction du bruit

On voit bien que les coordonnées planimétriques des segments reconstruits sont beaucoup plus fiables et robustes au bruit que les coordonnées altimétriques. On retrouve également le rapport entre les précisions altimétrique et planimétrique se traduisant par $\sigma_{xy} = \sigma_z * B/H$ (cf. section 2.2.4.1).

4.4.1.3 Précision en fonction du rapport B/H :

Les données dont on dispose en entrée ne présentent pas différents rapports B/H sur la même zone. Les évaluations en fonction du rapport B/H ne pourront donc être réalisées que par le biais de simulations en modifiant explicitement les coordonnées des centres de prise de vue. Cette manipulation est possible dans ce cas, puisqu'on n'utilise pas la radiométrie des images. La fourchette de B/H s'étend de 0.05 à 0.8. Ce sont les valeurs prévues en théorie pour les futurs satellites PLÉIADES. La figure 4.7 illustre séparément les précisions en XY et en Z ainsi que les écarts-types sur les mesures. On voit bien que plus le rapport B/H est grand, plus la reconstruction est précise. Dans le cadre de cette thèse, on s'intéresse à la fourchette [0.05 - 0.2]. Les simulations confirment le fait que la précision altimétrique est très dégradée, par contre on dispose d'une bonne localisation planimétrique des segments 3D. Les algorithmes présentés devront tirer profit de cet avantage.

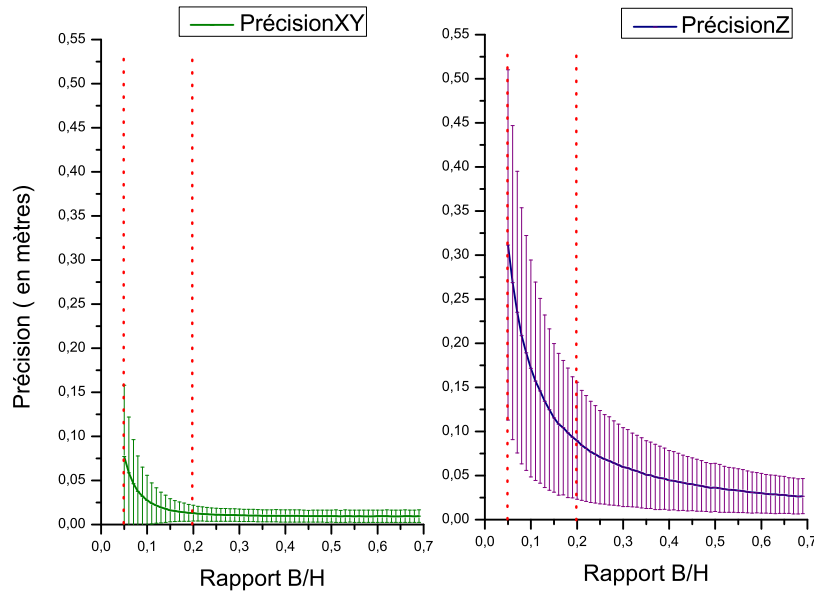


FIG. 4.7 – Précision de reconstruction en fonction du rapport B/H

4.4.2 Précision par rapport à un modèle de référence

Pour s'affranchir des problèmes de détection de segments dans les images, des simulations ont été réalisées en utilisant un modèle vecteur de référence. Afin de tester la pertinence de la bande de tolérance, il est nécessaire que le modèle vecteur soit cohérent avec le MNE dont on dispose en entrée, au niveau des localisations planimétriques des bâtiments et de leurs altitudes. D'autre part, afin d'estimer la précision de la reconstruction 3D, il est nécessaire de s'affranchir des

problèmes d'aérotriangulation. Par conséquent, on n'utilisera pas la référence absolue (TRAPU[®] de l'IGN) mais un modèle vecteur saisi manuellement par un opérateur humain à partir des images de simulations PLÉIADES.

La scène illustrée figure 4.8 servira comme référence à toutes les simulations. Les façades ont été supprimées pour éliminer les parties cachées lors des projections des segments dans les images.

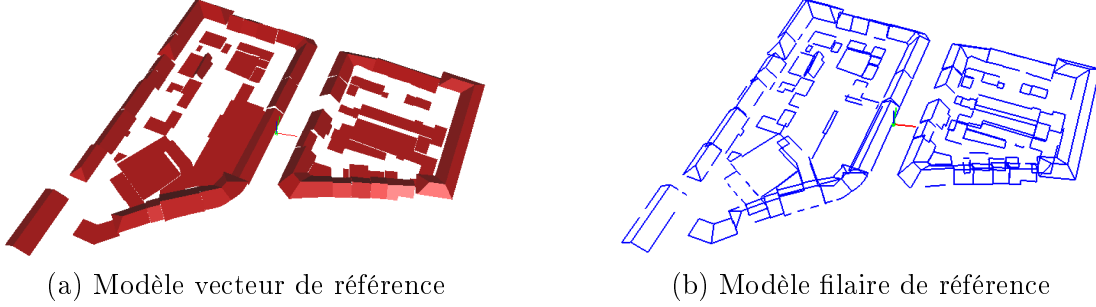


FIG. 4.8 – Modèle de référence pour les simulations

Le nombre de segments de la scène est de 546. Les segments 3D sont projetés dans les deux images. Les segments 2D sont échantillonnés à la résolution de l'image. Les points constitutifs des segments 2D sont bruités avec un bruit gaussien isotrope d'écart type σ . Une batterie de 1000 tirages est réalisée pour chaque valeur de bruit. Les segments 2D bruités sont ensuite appariés et reconstruits en 3D.

4.4.2.1 Exhaustivité de la reconstruction

La figure 4.9 illustre le pourcentage de segments 3D reconstruits, ce qui revient au nombre de segments 2D appariés en fonction du bruit. La figure reporte également le nombre de segments

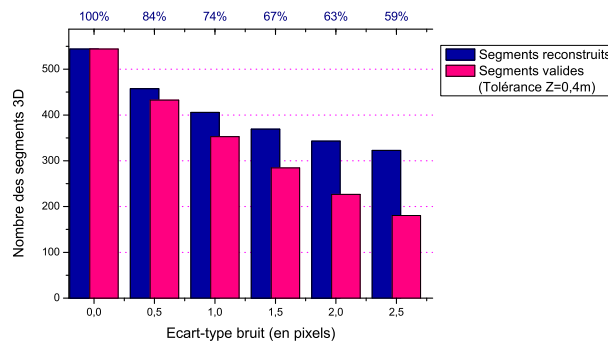


FIG. 4.9 – Exhaustivité de la reconstruction en fonction du bruit

3D valides. Un segment 3D est considéré comme **valide** si ses deux extrémités appartiennent à la bande de tolérance dérivée du MNE initial. Pour cet exemple, on a pris comme valeur de bande de tolérance $\sigma_Z = 0.4m$. On remarque qu'en absence de bruit, tous les segments sont reconstruits. En pratique, pour un $\sigma_{bruit} = 0.5$ pixel, 84% des segments sont reconstruits dont

94% qui sont valides. On remarque également que le nombre de segments 3D valides décroît plus vite que le nombre de segments reconstruits, c'est-à-dire que les associations de segments 2D peuvent être obtenues en présence du bruit, mais la reconstruction 3D n'est plus fiable. Ceci montre la nécessité d'une validation a posteriori des segments 3D reconstruits. L'apport de la bande de tolérance pour la validation des segments 3D sera démontré dans la section suivante.

4.4.2.2 Influence de la bande de tolérance :

La bande de tolérance est un volume dilaté autour du MNE de la valeur de la tolérance en Z , $\pm\sigma_z$ (cf. figure 4.3). Pour évaluer l'influence de la bande de tolérance sur la précision des segments retenus, deux écarts-types sont retenus pour le bruit ; $\sigma_{bruit} = 0.5$ et $\sigma_{bruit} = 0.1$ ce qui correspond pour une résolution de 50 cm à une erreur de mesure respective de 0.25 m et 0.5 m. La figure 4.10 illustre le nombre de segments valides en fonction de la tolérance en Z ainsi que l'évolution de la moyenne de précision sur ces segments. On remarque que plus la tolérance en Z est grande, plus le nombre de segments valides est grand, mais la moyenne de la précision se dégrade. Un compromis sera à trouver entre l'exhaustivité et la précision de la reconstruction.

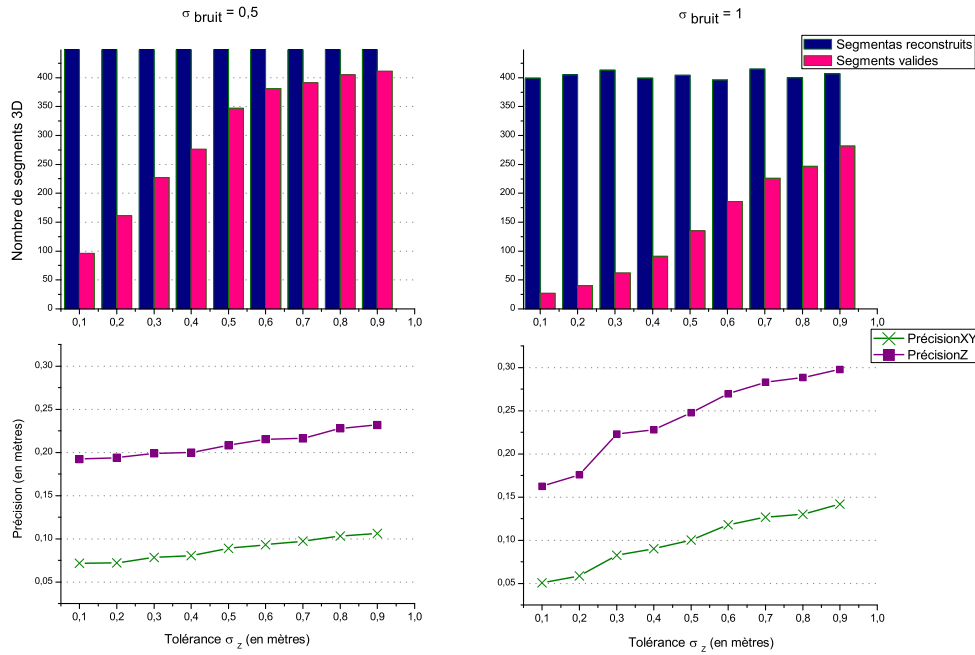


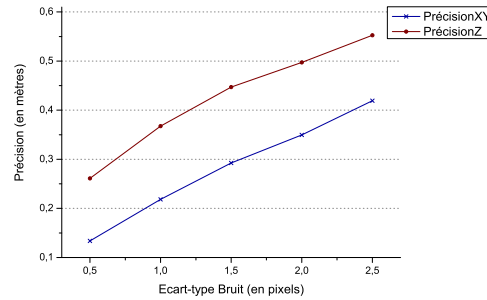
FIG. 4.10 – Précision en fonction de la bande de tolérance

Si l'application visée nécessite des segments fiables en altimétrie au détriment de l'exhaustivité, une faible valeur de Z sera choisie pour la bande de tolérance. Au contraire, une grande valeur de tolérance en Z favorise l'exhaustivité aux dépens de la fiabilité. On remarquera toutefois, pour ce dernier cas, que le manque de fiabilité concerne surtout les coordonnées altimétriques. Les coordonnées planimétriques restent fiables grâce au faible rapport B/H du couple.

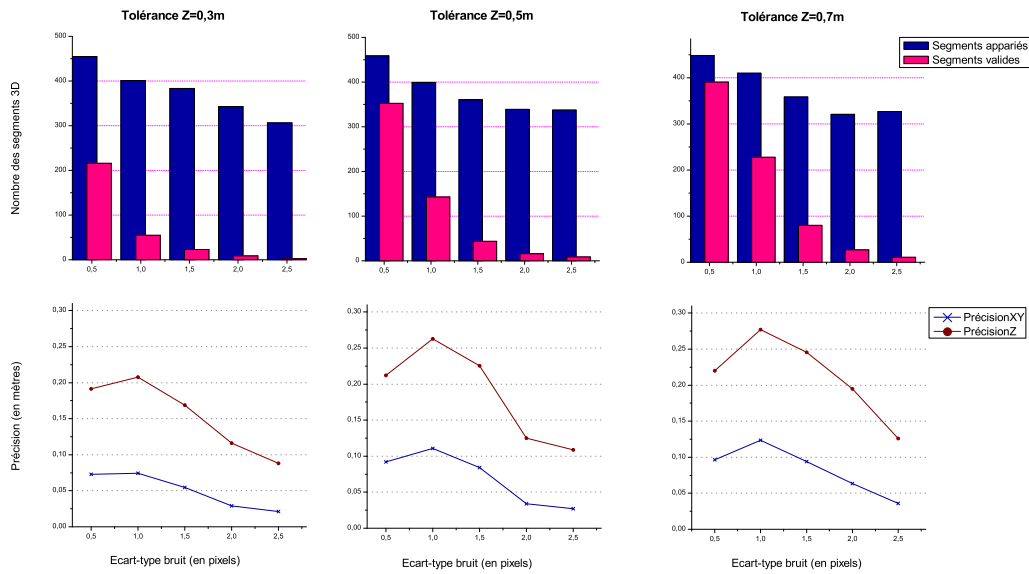
La section suivante illustre la précision des segments reconstruits en présence de bruit.

4.4.2.3 Précision en fonction du bruit des images :

Les figures 4.11(a) et 4.11(b) illustrent la moyenne de la précision de reconstruction sur l'ensemble des segments avec différentes valeurs d'écart-type pour le bruit.



(a) Sans bande de tolérance



(b) Validation avec une bande de tolérance

FIG. 4.11 – Simulations - Précision de reconstruction 3D en présence de bruit

La figure 4.11(a) illustre l'évolution de la précision des segments 3D reconstruits sans l'étape de validation par une bande de tolérance. La précision se dégrade avec le bruit. Dans la figure 4.11(b), une bande de tolérance est utilisée et trois valeurs de tolérance en Z sont comparées : $\sigma_Z = 0.3m$, $\sigma_Z = 0.5m$ et $\sigma_Z = 0.7m$. La moyenne de précision est calculée uniquement sur les segments valides. On voit clairement que la bande de tolérance améliore la précision de la

reconstruction 3D aux dépens du nombre de segments valides.

4.5 Résultats sur des images réelles

L'algorithme de reconstruction de segments 3D a été testé sur un couple d'images de simulation PLÉIADES sur Amiens à 50 cm et 70 cm. La taille de l'image à 50 cm est de 725* 856 pixels, ce qui correspond à une surface de 15,51 *ha*. Les paramètres utilisés sont :

- Suréchantillonnage de l'image : 2
- Alpha du filtre de canny Deriche : 1.5
- Seuillage hystérésis : $s_b = 0$, $s_h = 5$
- Tolérance Z : 0.4 m
- Seuil de corrélation = 0.5 / Fenêtre de corrélation = 5 pixels pour 50cm et 3 pixels pour 70cm
- Poids des contraintes de mise en correspondance : $\lambda_G = 0.4$ et $\lambda_{Ph} = 0.6$

4.5.1 Résultats qualitatifs

Pour des images réelles, l'évaluation quantitative de la précision des segments 3D reconstruits est plus difficile puisqu'elle nécessite une étape de mise en correspondance des segments 3D reconstruits et ceux de référence afin de les comparer. Par manque de temps, on se contentera, d'une évaluation qualitative par un opérateur humain.

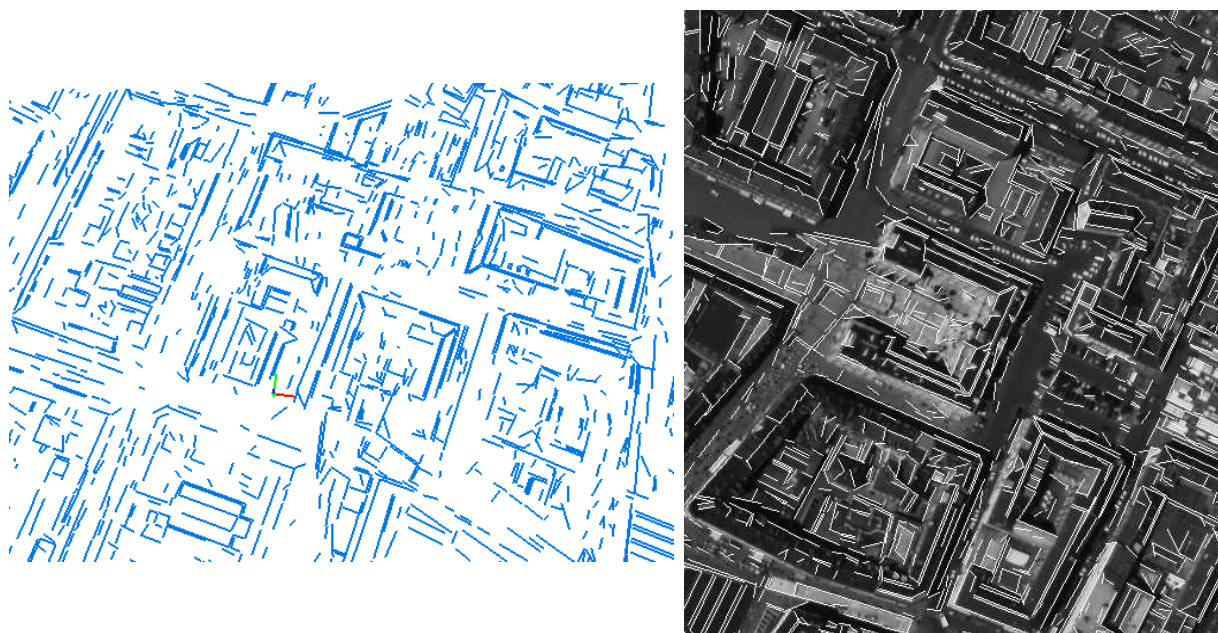
Les figures 4.12(a) et 4.12(b) illustrent les segments reconstruits sur la zone de test ainsi que leurs projections dans les images à différentes résolutions. On voit que les segments se reprojettent bien sur les contours de l'image, ce qui confirme la fiabilité des coordonnées planimétriques.

4.5.2 Impact des contraintes

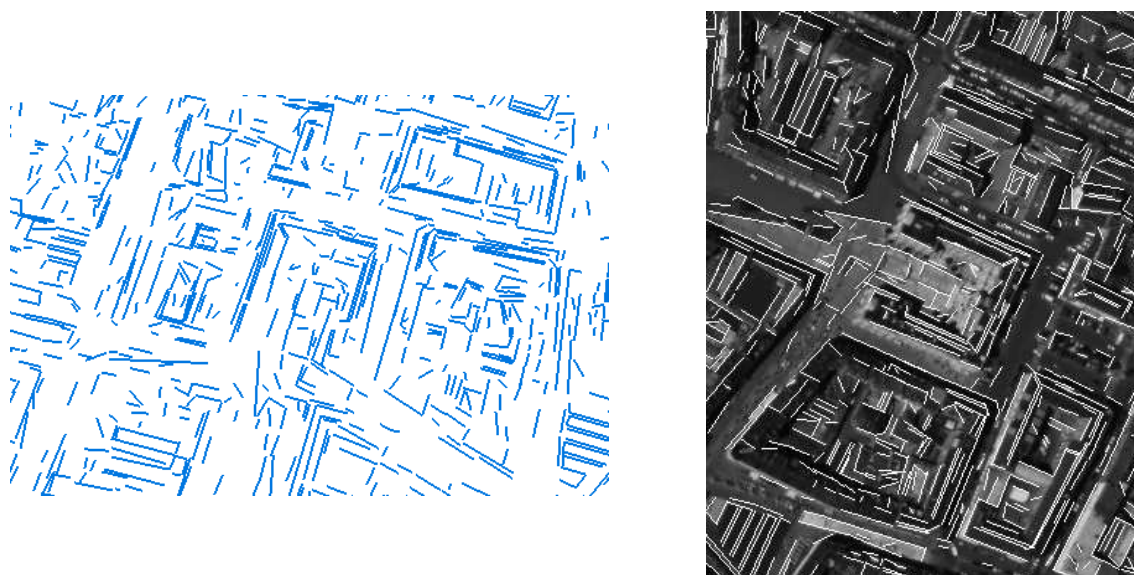
Le tableau 4.1 montre l'impact des différentes contraintes utilisées sur la reconstruction des segments 3D. La même zone test a été utilisée à 50 et 70 cm. Le tableau reproduit toute la chaîne de traitement. Tout d'abord, les contraintes de mise en correspondance sont appliquées indépendamment l'une de l'autre. Les contraintes d'élagage sont appliquées en série, dans l'ordre indiqué et le nombre d'associations retenues pour chaque contrainte tient compte des contraintes précédentes. Les images sont traitées de manière symétrique. Dans la colonne image 1 par exemple, l'image 1 est considérée comme image de référence pour le processus de mise en correspondance.

TAB. 4.1 – Impact des contraintes pour la détection des segments 3D

	Simulations Amiens	50cm		70cm	
	Image	1	2	1	2
	Nb segments 2D	7345	7574	3715	3768
Contraintes de mise en correspondance	Contrainte de recouvrement	6121	6259	3028	3102
		Moy Assocs = 2.72		Moy Assocs = 2.67	
	Contrainte photométrique	6656	6602	4611	4606
		Moy Assocs = 2.10		Moy Assocs = 3.30	
Contraintes d'élagage	Bande de tolérance	3345 (45%)	3340 (44%)	1460 (39%)	1462 (39%)
	Contrainte unicité	2883 (38%)		1286 (34%)	
	Contrainte orientation	2524 (34%)		1136 (30%)	



(a) Amiens 50 cm



(b) Amiens 70 cm

FIG. 4.12 – Segments 3D reconstruits et leurs projections dans les images

Les pourcentages présentés dans le tableau sont calculés par rapport au nombre de segments 2D extraits dans l'image.

Les deux contraintes de mise en correspondance sont complémentaires. La contrainte photométrique permet de lever l'ambiguïté entre deux segments renvoyant le même score de recouvrement, tel est le cas des segments de faîtes de toits qui sont en général dédoublés, ou bien de pallier

aux problèmes de sous-segmentations de segments entre les deux images et qui renvoient de faibles scores de recouvrement. Cependant, la contrainte photométrique a tendance à pénaliser les segments aux bords de bâtiments à cause de la mauvaise corrélation le long des façades.

Le tableau 4.1 montre également l'importance de la bande de tolérance pour élaguer les segments 3D non valides. Seulement 45% des segments appartiennent à la bande de tolérance à 50 cm, 39% des segments sont valides à 70 cm.

On constate d'après ce tableau, la difficulté de reconstruire de manière satisfaisante des paysages complexes à partir d'images satellitaires très haute résolution : si une résolution de 50 cm permet encore de récupérer un nombre suffisant d'appariements précis, la résolution à 70 cm donne des résultats nettement moins satisfaisants (dans un rapport de 2 à 1 environ).

4.6 Conclusions

On a présenté dans ce chapitre un algorithme d'extraction de segments 3D à partir d'un couple d'images satellitaires très haute résolution. L'algorithme a l'avantage de traiter les deux images de façon symétrique, mais aussi d'utiliser la contrainte photométrique afin de mettre à profit le faible rapport B/H du couple. Comme stratégie d'appariement, on a choisi de combiner les contraintes géométriques et photométrique. Comme stratégie de validation, des contraintes d'unicité et de validation par rapport à une bande de tolérance autour du MNE initial ont été utilisées.

L'évaluation théorique à partir de simulations a permis de montrer les limitations de la précision de reconstruction en fonction de l'orientation des primitives linéiques par rapport à la base des centres de prises de vue et en fonction du rapport B/H .

Les simulations ont également relevé le compromis à trouver entre l'exhaustivité de la reconstruction et la fiabilité des segments 3D. On privilégiera l'une ou l'autre en fonction de l'application. Les segments 3D extraits peuvent être utilisés pour extraire des plans 3D pivotants (cf. chapitre 5). Dans ce cas, on ne gardera que les segments fiables aussi bien en altimétrie qu'en planimétrie aux dépens de l'exhaustivité d'extraction, sinon l'erreur altimétrique sur les segments sera propagée dans la suite du processus d'extraction des plans 3D.

Les segments 3D seront également utilisés pour contraindre l'optimisation globale afin d'extraire la surface 3D finale. Les segments 3D modéliseront les discontinuités. Dans ce cas, la condition sur les segments extraits sera uniquement la bonne localisation planimétrique et on privilégiera par conséquent l'exhaustivité dans ce cas.

Un inconvénient de cette méthode est l'absence d'un critère de fiabilité pour les segments 3D reconstruits.

4.7 Perspectives

Dans notre algorithme, on a opté pour une mise en correspondance des segments dans l'espace image. Dans la perspective de la reconstruction de segments 3D à partir de triplets d'images PLÉIADES, il sera plus judicieux de se placer dans l'espace objet [Tai04].

Les segments 3D constituent des primitives linéiques fiables en planimétrie. Dans la suite de l'étude, on proposera des algorithmes de reconstruction de surfaces planes qui constituent des primitives complémentaires aux segments 3D. Dans le processus d'optimisation globale, les segments permettront de modéliser les discontinuités, les facettes 3D permettront de régulariser les surfaces 3D.

Chapitre 5

Reconstruction de facettes 3D par approche segments

5.1 Introduction

L'objectif final de cette étude est de proposer, pour des scènes urbaines, une modélisation de surfaces 3D avec plusieurs niveaux d'interprétation suivant la fiabilité des primitives 3D extraites. Dans le chapitre 4, on a proposé un algorithme d'extraction de primitives linéiques. On s'intéresse dans la suite à l'extraction de primitives surfaciques. On cherche alors à extraire des facettes 3D fiables qui seront utilisées dans un processus d'optimisation globale. Deux méthodologies seront proposées. Une première approche basée sur les segments 3D extraits sera développée dans ce chapitre. Dans le chapitre 6, une deuxième approche basée sur les régions sera détaillée.

5.2 Méthodologie

Dans la première approche, l'extraction de facettes 3D est basée sur les segments 3D. Le principe est illustré dans la figure 5.1. Les segments 3D sont obtenus par mise en correspondance des segments 2D dans les deux images. Ensuite, l'extraction des plans 3D revient à rechercher des demi-plans pivotant autour des segments 3D extraits. Finalement, un algorithme de délimitation des toits sur les plans porteurs permettra d'obtenir les facettes 3D finales.

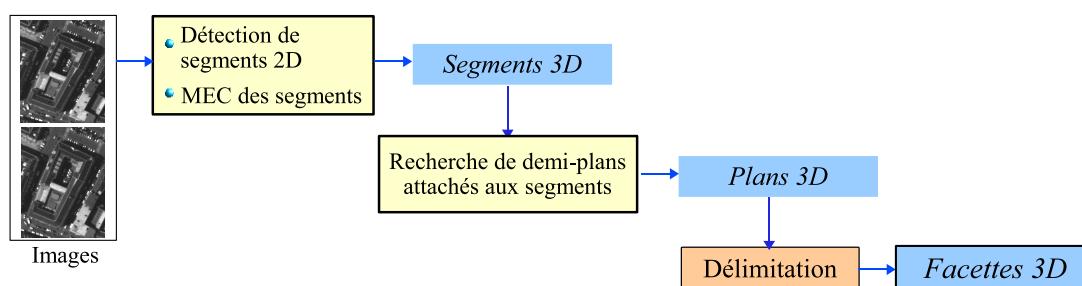


FIG. 5.1 – Extraction de facettes 3D : approche segments

5.3 Extraction des plans 3D

Cette méthode essentiellement dérivée de [BZ99] a été adaptée à notre contexte. La méthode consiste à extraire des demi-plans fiables définis uniquement par un segment 3D et les scores de similarité de son voisinage entre les deux vues. Cette méthode a été retenue puisqu'elle utilise les deux contraintes géométrique et photométrique. L'objectif est de déterminer pour chaque segment 3D, à travers deux familles de 1-paramètre θ ; $\Pi(\theta)$ et $\Pi(\theta')$ s'il y a deux demi-plans attachés ou pas. Si les demi-plans existent, on cherche la meilleure estimation de θ et θ' . Les deux familles représentent les deux demi-plans attachés au segment de part et d'autre de l'axe vertical. Tout d'abord, on extrait un ensemble de points texturés au voisinage du segment qui correspondent à des points d'intérêt dans l'image [HS88] (cf. figure 5.2).

Pour avoir une fonction de similarité discriminante, la corrélation sera calculée au voisinage de ces points d'intérêt *POI*. Étant donnée une hypothèse angulaire de θ , le plan $\Pi(\theta)$ définit une correspondance pixel à pixel entre les deux images. Si le plan est correct, le score de corrélation entre les points correspondants doit être élevé. L'hypothèse d'angle θ qui sera retenue est celle qui renvoie le maximum de la fonction de similarité. Cette fonction de similarité $Sim(\theta)$ doit être sélective et robuste aux points aberrants. Elle est définie comme suit et varie entre $[0, 1]$.

$$\begin{cases} \theta_{plan} = \underset{\theta}{argmax}(Sim(\theta)) \\ Sim(\theta) = \frac{1}{2} \sum_{i=1,2} \iint_{POI(img_i)} Corr^2(x, x') \end{cases} \quad (5.1)$$

Le terme de corrélation $Corr(x, x')$ correspond à la corrélation centrée normalisée entre le pixel x dans une image et son pixel homologue x' dans la seconde image. Le terme est mis au carré pour donner plus de poids aux scores élevés et pour rendre la fonction de similarité plus sélective.

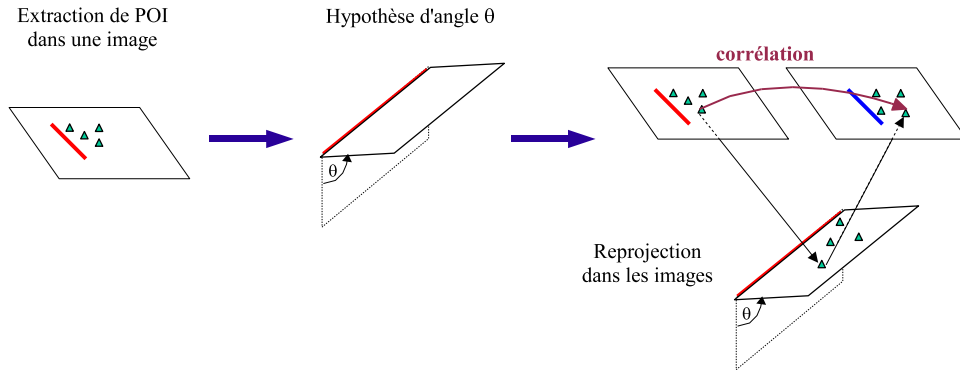


FIG. 5.2 – Extraction des demi-plans pivotants

La corrélation est discriminante sur des régions texturées d'où l'intérêt d'utiliser des *POI*. Dans notre cas, les *POI* sont calculés dans les deux images et traités de façon symétrique donc le score de corrélation final pour un plan donné est moyenné sur les deux vues (cf. eq 5.1). Ceci augmente la fiabilité des plans 3D détectés.

On utilise pour la corrélation, des fenêtres adaptatives dont le masque est déterminé par les contours extraits. L'angle optimal θ renvoie le maximum de la fonction de similarité $Sim(\theta)$ avec $\theta \in [20^\circ, 160^\circ]$ par rapport à une référence verticale. Le processus est itéré pour θ' afin d'extraire le deuxième demi-plan. On ne garde que les demi-plans dont le score est supérieur à un seuil prefixé ($S_{th} = 0.5$).

5.3.0.1 Contrainte de cohérence des demi-plans

On suppose que les demi-plans sont orientés. Les orientations de deux demi-plans rattachés au même segment, c'est à dire adjacents, sont cohérentes si elles induisent sur l'arête commune, des orientations opposées (cf. figure 5.3). Ce qui se traduit par la condition suivante $[s, v_1, n_1] \cdot [s, v_2, n_2] < 0$ [Tai04].

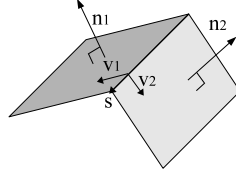


FIG. 5.3 – Cohérence des demi-plans adjacents

5.4 Délimitation des facettes 3D

5.4.1 État de l'art

L'étape précédente a permis de déterminer les positions angulaires des demi-plans qui vont servir de support aux pans de toits. L'algorithme présenté ici, a pour but de délimiter les facettes 3D, ce qui revient à déterminer l'étendue de chaque pan de toit, de manière à travailler non plus sur des demis plans, mais avec des polygones plans positionnés dans l'espace 3D. Le problème de délimitation de facettes est délicat. Dans la littérature [AF99, VD01], les auteurs utilisent en général une étape de fusion et d'intersection des plans 3D détectés pour délimiter les facettes 3D. Cette étape présente une grande complexité en fonction du nombre de plans 3D présents dans la scène. Les intersection de plans ont surtout lieu au niveau des faîtes de toits et des coins de bâtiments. En effet, étant donnée, l'absence de plans verticaux liés aux façades, difficilement détectés en stéréoscopie, l'étape d'intersection de plans ne permet pas de délimiter les facettes sur les bords de bâtiments. Dans [BZ99], les plans 3D attachés à des segments sont délimités par d'autres segments 3D dans leur voisinage. D'autres travaux disposent de données externes qui fournissent les limites planimétriques des bâtiments et permettent de délimiter les toits telles que les limites cadastrales [JPDP00].

Dans notre contexte, les segments 3D extraits correspondent en partie aux faîtes de toits. Pour délimiter les toits, on propose un algorithme basé sur la relaxation probabiliste qui permet de propager l'information fiable au voisinage des segments 3D et qui est surtout basé sur les scores de corrélation permettant de mettre à profit le faible rapport B/H du couple.

5.4.2 Délimitation des toits par relaxation probabiliste

5.4.2.1 Principe

Tout d'abord, la relaxation probabiliste [FY97] s'effectue sur un graphe, c'est-à-dire un ensemble d'objets appelés sommets ayant des caractérisations communes et des relations entre eux matérialisées par des arêtes.

L'idée fondamentale des méthodes de relaxation sur un graphe est de faire évoluer de façon itérative des caractéristiques de chacun des sommets en fonction des caractéristiques de ses voisins, de manière à propager l'information dans le graphe, et ce jusqu'à une position d'équilibre.

Dans le cas de la relaxation probabiliste, les sommets sont caractérisés entre autres par des états probabilisés et l'on fait évoluer ces probabilités. La relaxation probabiliste a été utilisée au sens de Rosenfeld ; on a donc la relation itérative suivante pour la probabilité à l'itération $k + 1$ que le sommet i soit dans l'état λ :

$$p_i^{(k+1)}(\lambda) = p_i^{(k)}(\lambda) \cdot \frac{1 + s_i^{(k)}(\lambda)}{1 + \bar{s}_i^{(k)}} \quad (5.2)$$

et

$$s_i^{(k)}(\lambda) = \sum_{j=1}^n d_{ij} \sum_{\lambda'=1}^m c_{ij}(\lambda, \lambda') p_i^{(k)}(\lambda') \quad (5.3)$$

$$\bar{s}_i^{(k)}(\lambda) = \sum_{\lambda=1}^m p_i^{(k)}(\lambda) s_i^{(k)}(\lambda) \quad (5.4)$$

avec

- Un ensemble de m labels : $\{\lambda_i, i = 1, \dots, m\}$ pour chaque région.
- $c_{ij}(\lambda, \lambda')$ représente la mesure de compatibilité que le sommet i ait le label λ quand le sommet j a le label λ' . Cette compatibilité peut être considérée comme une probabilité conditionnelle qui satisfait la condition suivante :

$$\begin{cases} \forall \lambda, \lambda' \ 0 \leq c_{i,j}(\lambda, \lambda') \leq 1. \\ \sum_{\lambda} c_{i,j}(\lambda, \lambda') = 1 \end{cases} \quad (5.5)$$

- d_{ij} quantifie l'influence du sommet i sur le sommet j . Elle vérifie $\sum_j d_{ij} = 1$.

Ainsi, lors d'une relaxation probabiliste, trois paramètres sont à régler :

- La fonction d'influence d_{ij}
- Les probabilités conditionnelles c_{ij}
- L'ensemble des probabilités initiales dénotées $p_i^0(\lambda)$, $\lambda = 1, \dots, m$.

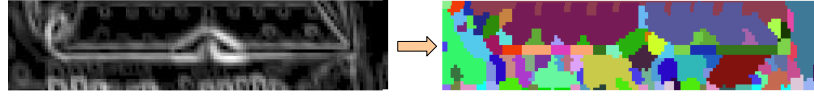
5.4.3 Application à notre problème

Étant donné que le demi-plan est positionné, on travaillera, dans ce qui suit, sur des imageries issues de la reprojection des abords du segment considéré sur le plan $\Pi(\theta)$ dans l'espace 3D [Alc01]. Dans un premier temps, une accumulation des gradients des deux imageries du couple stéréoscopique est réalisée. Dans notre cas, la relaxation probabiliste est appliquée à un graphe d'adjacence, c'est-à-dire un graphe où les sommets sont des régions et les arêtes indiquent une frontière commune entre deux régions (cf. figure 5.4-b). L'image des régions est obtenue par une sur-segmentation réalisée par l'algorithme de partage des eaux [CP97] sur l'image d'accumulation des gradients (cf. figure 5.4-a).

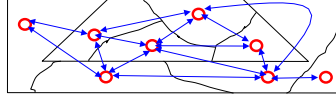
On définit ensuite pour ce graphe les états probabilisés, qui sont au nombre de deux : " \in toit" et " \notin toit". Le réglage des trois paramètres de la relaxation probabiliste est détaillé dans ce qui suit :

5.4.3.1 Les probabilités initiales

Les probabilités initiales d'appartenance au toit ont été fixées à 99% pour les régions adjacentes au segment et à la valeur moyenne de la fonction de similarité sur la région dans les



(a) Accumulation de gradients + watershed



(b) Graphe d'adjacence des régions

FIG. 5.4 – Construction du graphe d'adjacence

autres cas. La fonction de similarité correspond au score de corrélation entre les deux imagerie reprojctées sur le plan. Le score de corrélation est pondéré de $\frac{1}{d}$, où d est la distance du barycentre de la région au segment. En effet, les régions adjacentes au segment ont de fortes chances d'appartenir au toit, et le fait d'avoir une zone de forte probabilité permet de propager cette information aux régions voisines qui sont susceptibles d'appartenir au toit.

D'autre part, le score de similitude est lui aussi significatif de l'appartenance au toit, car dans les zones en dehors du toit, la reprojction dans l'espace objet n'est pas pertinente et il y a donc des déformations, d'où un score de corrélation faible, alors que les zones contenues dans le toit sont reprojctées correctement, ce qui donne un score de similitude élevé.

5.4.3.2 Les probabilités conditionnelles

Le choix des probabilités conditionnelles n'est pas critique, on choisit de les fixer de la manière suivante :

$$\begin{cases} Labels : \lambda, \lambda' \in \{ " \in toit", " \notin toit" \} \\ c_{ij}(\lambda, \lambda) = 0.6 \\ c_{ij}(\lambda, \lambda') = 0.4 \end{cases} \quad (5.6)$$

Les probabilités conditionnelles sont choisies aussi proches que possible de 0,5 de manière à favoriser une évolution lente des probabilités.

5.4.3.3 La fonction d'influence

La fonction d'influence d_{ij} est un paramètre critique. Elle doit satisfaire un certain nombre de conditions préfixées :

- être croissante en fonction de la longueur de la frontière commune entre les deux régions ;
- être décroissante en fonction de la valeur moyenne du gradient le long de la frontière commune aux deux régions. En effet, plus le contour est marqué entre deux régions, plus il y a de chance qu'elles n'appartiennent pas au même objet dans l'image ;
- le gradient doit l'emporter sur la longueur commune.

D'après les conditions citées ci-dessus, on propose la fonction d'influence suivante [Alc01] :

$$d_{ij} = \frac{L_{ij} \cdot e^{-\lambda \langle grad \rangle_{ij}}}{\sum_{k=0}^n L_{ik} \cdot e^{-\lambda \langle grad \rangle_{ik}}} \quad (5.7)$$

où

- L_{ij} est la longueur de la frontière commune aux deux régions.
- $\langle grad \rangle_{ij}$ est la valeur moyenne du gradient sur la frontière commune aux régions i et j .
- On utilise $\lambda = 3.5$. Cette valeur est choisie de manière empirique. Elle permet de délimiter les toits principaux de la scène. Dans les cas litigieux, elle tend à renvoyer des toits plus petits qu'en réalité, ce qui est préférable pour la lisibilité de la scène 3D générée.

L'évolution du système conduit à une carte finale des probabilités (cf. figure 5.5). Les régions les plus claires ont de fortes probabilités d'appartenir au toit. Finalement, une coupe dans le graphe d'adjacence permet de sélectionner les régions connexes qui ont une probabilité d'appartenir au toit supérieure à un seuil fixé $P_s = 0.75$. La plus grande région connexe sera choisie comme délimitation finale du toit.

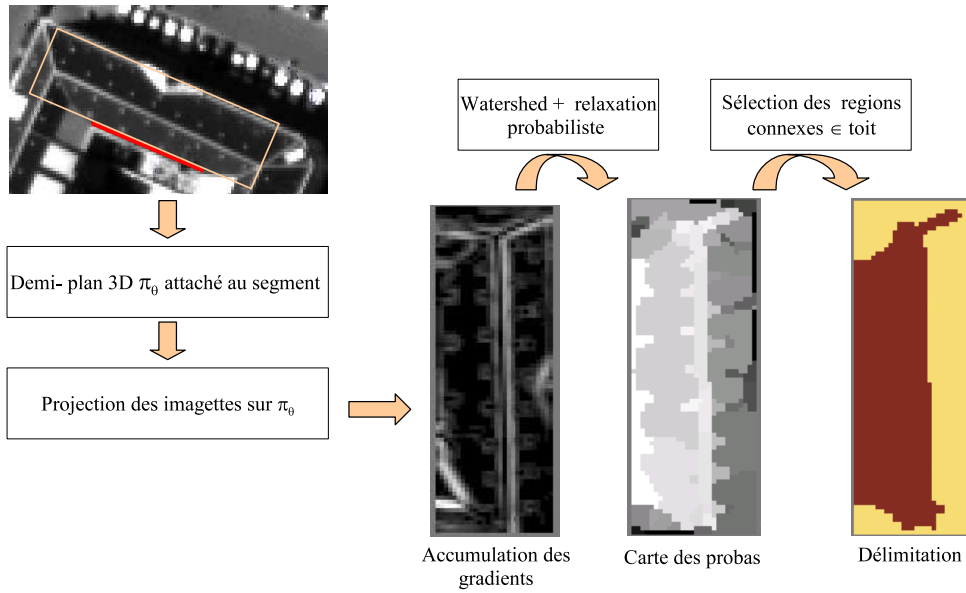


FIG. 5.5 – Délimitation des facettes 3D par relaxation probabiliste

5.5 Résultats & Évaluations

Les tests ont été réalisés sur un couple d'images de simulation Pléiades sur Amiens à 50 et 70 cm avec un rapport $B/H = 0.2$. Les évaluations sont présentées sur différentes zones.

5.5.1 Extraction des demis-plans

La figure 5.6 illustre un exemple de plans 3D extraits sur le centre ville d'Amiens. Les paramètres utilisés sont :

- Fenêtre pour le détecteur d'Harris 3*3
- Fenêtre de corrélation 5*5, Seuil de corrélation = 0.5
- Largeur imagerie : 5m

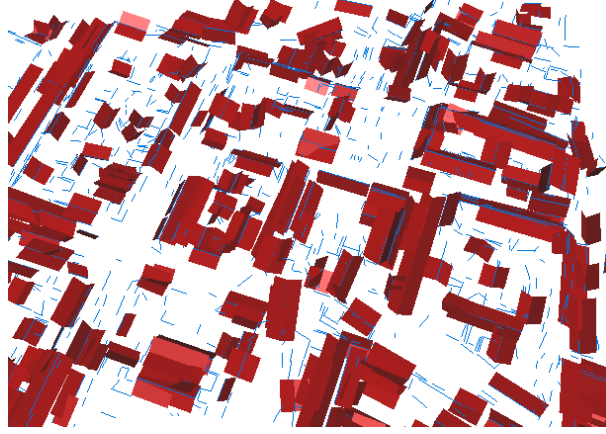
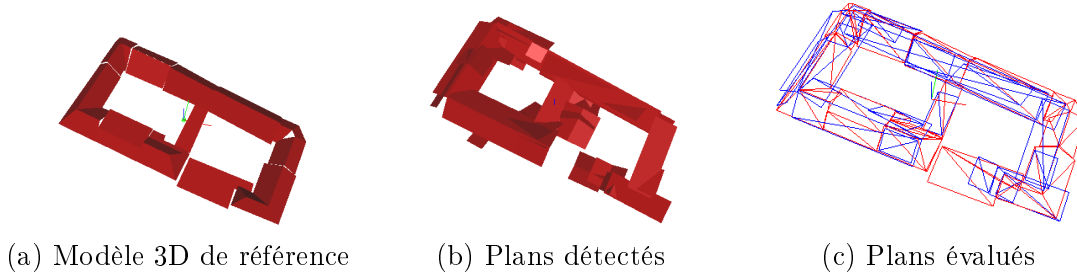


FIG. 5.6 – Extraction de plans 3D - Amiens 50cm

Les plans 3D ont été évalués grâce à la plateforme EvalMne[©] (cf. chapitre 8). Le modèle 3D de référence a été saisi manuellement par un opérateur humain à partir des images à 50 cm. La plateforme permet de réaliser la mise en correspondance automatique entre les facettes calculées et les facettes de référence (cf. section 8.3.4).

La figure 5.7 montre un exemple de plans 3D de référence et l'ensemble des plans détectés sur la zone. La figure 5.7-c illustre en modèle filaire la superposition des facettes appariées entre les deux ensembles. La référence est illustrée en rouge et les facettes évaluées en bleu.

FIG. 5.7 – Évaluation des plans par EvalMNE[©]

L'évaluation des plans portera sur l'exhaustivité et la précision de la reconstruction. Pour quantifier l'exhaustivité de la reconstruction, on introduit le taux de détection et le taux de sur-détection par rapport aux plans de référence que l'on notera T_d et T_{sd} (cf. section 8.3.1).

Pour mesurer la précision des plans 3D, on utilisera l'erreur angulaire ϵ entre deux plans qui se traduit par la valeur de l'angle entre les normales respectives aux plans. Les tableaux d'évaluations fournissent la moyenne de l'erreur angulaire et l'écart-type. Les facettes n'étant pas encore délimitées, on n'utilisera pas des critères d'évaluation sur la surface. Les évaluations sont présentées par bâtiments sur la figure 5.8. Le nombre de facettes correspond aux facettes appariées entre le modèle de référence et les plans détectés et qui sont effectivement évaluées.

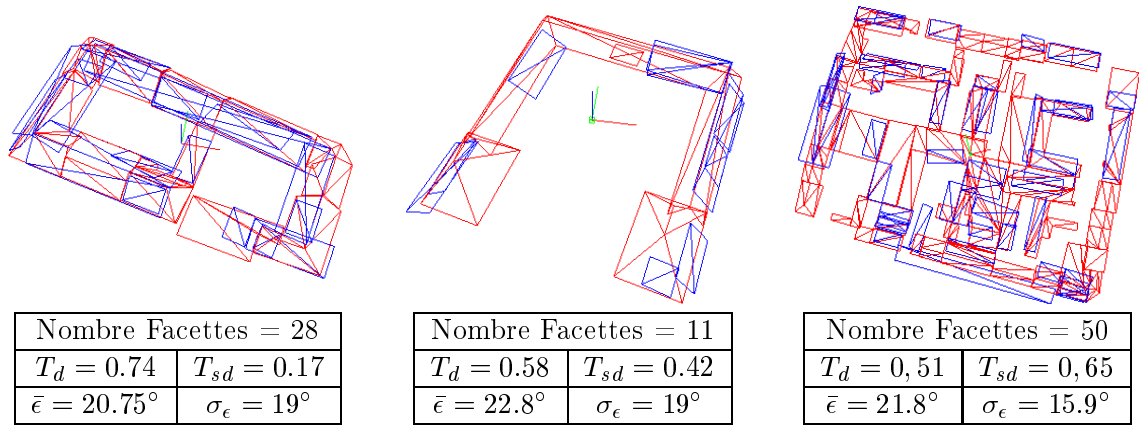


FIG. 5.8 – Evaluations de plans 3D par bâtiment

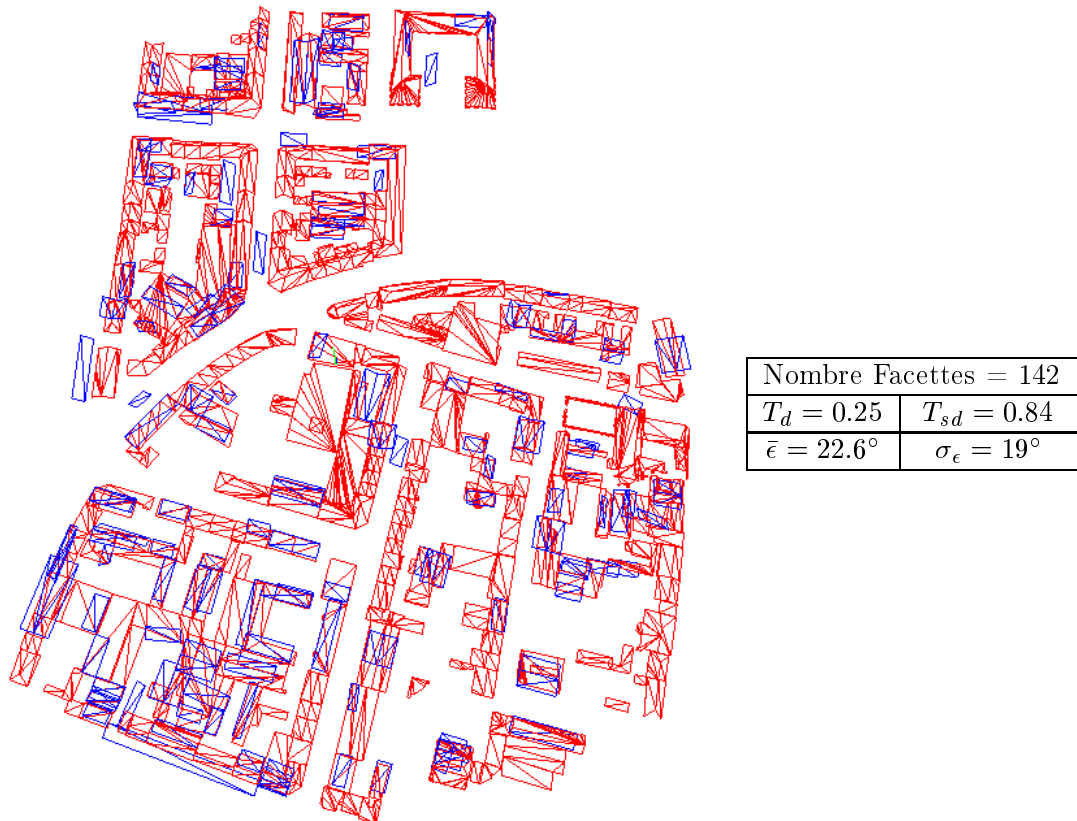


FIG. 5.9 – Exemple centre-ville d'Amiens 50 cm

On relève tout d'abord que le taux de détection est très variable et dépend essentiellement de la densité des segments 3D en entrée. Si des segments 3D ne sont pas détectés, des plans entiers seront manquants.

Sur les différentes zones de test, la moyenne de l'erreur angulaire est de 22° . Les erreurs sont principalement dues à des façades. En effet, les imagerie reprojetees sur les façades sont très déformées et par conséquent les scores de corrélation ne sont pas fiables. Ce qui fausse l'estimation de l'angle θ .

On remarque également, sur l'ensemble des zones, un fort taux de surdétection des plans 3D. Cet inconvénient était prévisible puisqu'on cherche des demi-plans rattachés à des segments 3D. Un demi-plan sera détecté, au mieux, autant de fois que le nombre de segments qui lui appartiennent. La surdétection des plans peut provenir aussi de la sursegmentation des segments initiaux. Ce problème peut être résolu par une étape de fusion des plans 3D a posteriori.

5.5.1.1 Évaluation en fonction de la résolution

La même zone de test a été utilisée à 50 cm et 70 cm. La figure 5.10 montre un comparatif sur un bâtiment en particulier. On remarque que le taux de détection des facettes 3D à 70 cm est nettement inférieur à celui à 50 cm. Par contre, la précision de la reconstruction est comparable en termes d'erreur angulaire par rapport aux plans de référence.

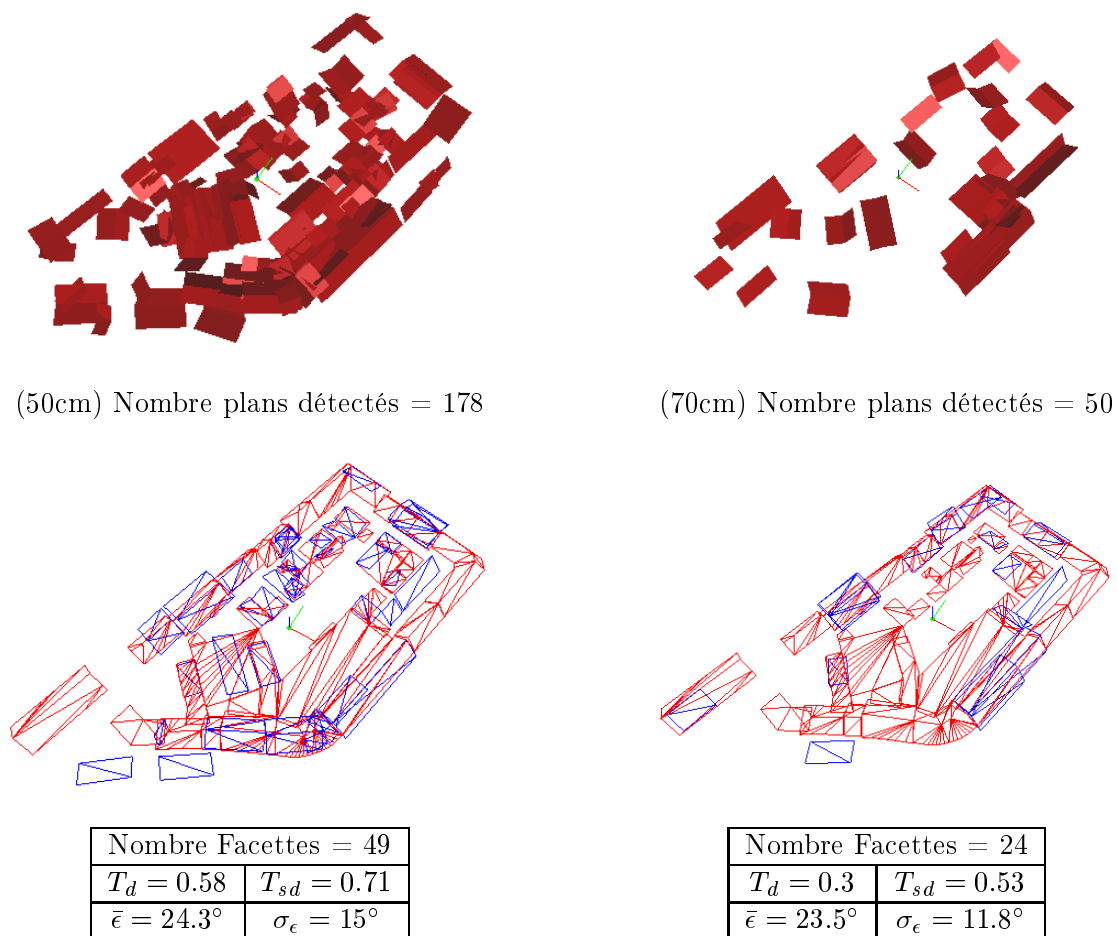


FIG. 5.10 – Evaluations de plans 3D en fonction de la résolution

5.5.1.2 Bilan

Les évaluations des plans 3D montrent en général un taux de surdétection élevé des facettes. L'exhaustivité de la reconstruction dépend de la densité des segments 3D en entrée du processus. La précision des plans 3D dépend également de la précision des segments 3D. Par ce fait, il est très difficile de qualifier les plans 3D extraits et de leur affecter des scores de fiabilité.

De plus, les plans sont détectés indépendamment les uns des autres. Aucune information topologique ne permet de valider ou invalider un plan 3D.

5.5.2 Délimitation des facettes 3D

Les figures 5.11 et 5.12 montrent des exemples de délimitation de facettes pour différents toits. La figure 5.13 montre la délimitation de facettes 3D sur le centre-ville d'Amiens à 50 cm. Pour des raisons de lisibilité des figures, la carte des probabilités est illustrée avec la palette RGB. Les régions n'appartenant pas au toit ayant une faible probabilité, sont illustrées en bleu.

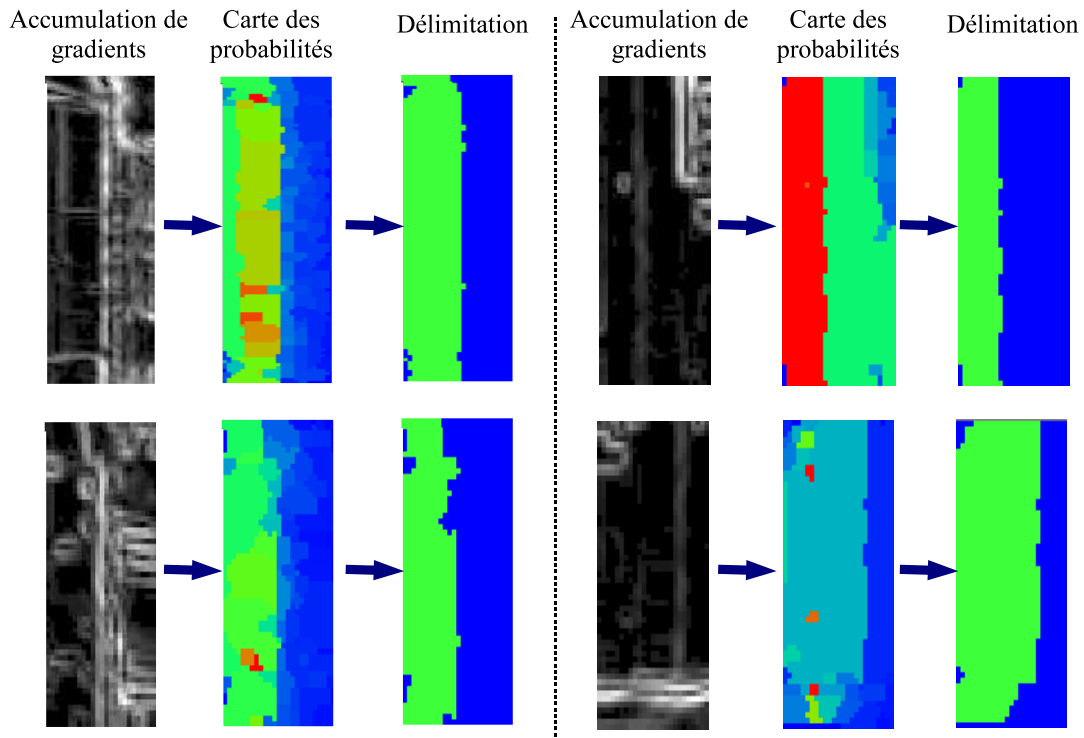


FIG. 5.11 – Exemples de délimitation de facettes 3D

La délimitation des facettes n'a de sens que pour des plans 3D fiables et correctement positionnés. De ce fait, l'évaluation portera uniquement sur les plans fiables. La délimitation est évaluée visuellement par un opérateur. Sur 200 plans fiables, la délimitation est correcte dans 68% des cas. Les erreurs de délimitation sont essentiellement dues aux segments dont les longueurs sont surestimées au départ (à cause de la stratégie d'union lors de la reconstruction 3D (cf. section 4.3.3)). Dans ce cas, l'initialisation des probabilités peut être faussée en considérant les régions proches du segment comme appartenant au toit alors qu'elles ne le sont pas en réalité.

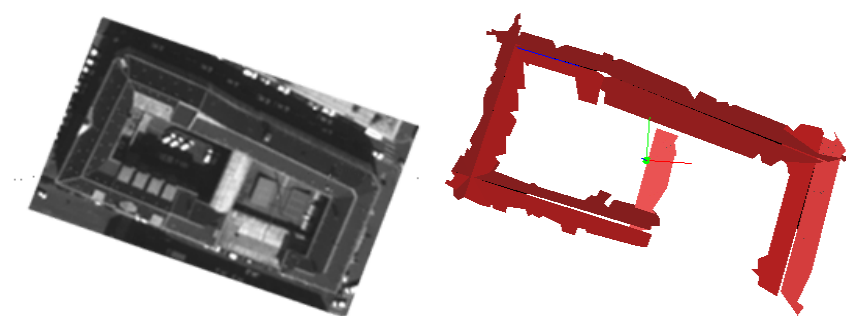


FIG. 5.12 – Amiens (France) - 50cm)

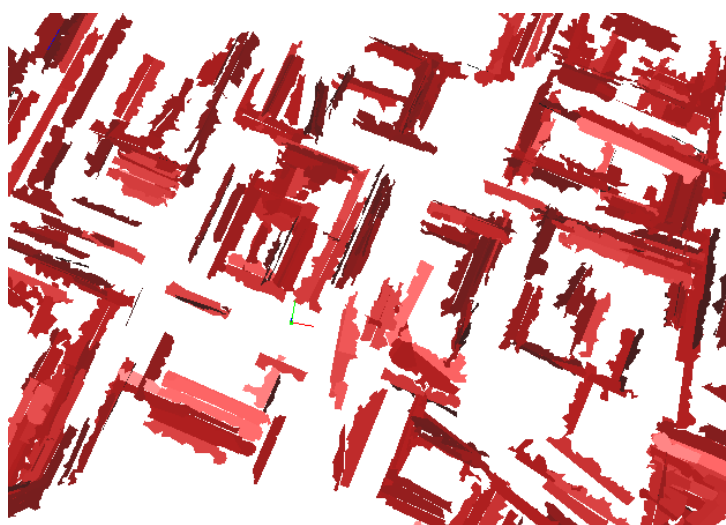


FIG. 5.13 – Exemple de délimitation des plans 3D, Amiens 50 cm

5.6 Conclusions

Ce chapitre détaille un algorithme d'extraction de facettes 3D basé sur les segments 3D. On distingue deux étapes dans le processus d'extraction des facettes 3D : la recherche des demi-plans 3D et la délimitation des facettes 3D.

5.6.1 Avantages

Les algorithmes ont l'avantage de traiter les deux images de façon symétrique, mais aussi d'utiliser la contrainte photométrique à chaque étape afin de mettre à profit le faible B/H du couple. Les deux algorithmes constituent les maillons d'une chaîne globale de reconstruction de facettes 3D, mais ils peuvent être utilisés séparément. Par exemple, l'algorithme de délimitation peut être utilisé pour délimiter a priori les surfaces de corrélation pour la recherche de demi-plans pivotants au lieu d'utiliser des surfaces rectangulaires.

5.6.2 Inconvénients

L'inconvénient de cette approche est le fait que les erreurs altimétriques sont propagées à chaque étape. En effet, la qualité de la reconstruction des plans 3D et la délimitation des facettes dépendent de la précision du segment 3D en entrée. Le taux de détection des plans 3D dépend également de celui des segments 3D.

D'autre part, la méthode souffre d'un fort taux de surdétection des plans qui est du d'une part à la méthodologie adoptée qui consiste à rechercher des demi-plans rattachés aux segments et d'autre part à la sursegmentation initiale des segments. Ce problème peut être résolu par une étape de fusion des plans a posteriori.

Le dernier inconvénient et sûrement le plus critique est la difficulté de qualifier les plans 3D extraits. En effet, leur extraction étant basée sur la maximisation des scores de corrélation, il est difficile d'utiliser le même critère pour les qualifier. De plus aucune information topologique ne permet de valider ou invalider les hypothèses de plans. Leur utilisation sera critique dans le processus d'optimisation globale puisqu'il nécessite des facettes fiables.

5.6.3 Bilan

Ces algorithmes permettent d'obtenir une modélisation de la scène qui n'est pas dense et entachée d'erreurs. Il est très difficile à partir des résultats d'obtenir un modèle polyédrique de bâtiment. Néanmoins, des applications potentielles existent. Ces primitives peuvent être utilisées dans les approches structurelles pour la reconstruction de bâtiments par exemple [FL99] où l'on cherche à construire un graphe de primitives à partir des données, pour l'apparier au mieux à une bibliothèque de graphes de modèles. Ces primitives peuvent également être utilisées comme contraintes dans les algorithmes d'extraction de surfaces.

Chapitre 6

Reconstruction de facettes 3D par approche régions

6.1 Introduction

Le but de ce chapitre est d'extraire des facettes 3D fiables qui seront utilisées dans un processus d'optimisation globale. L'ensemble des facettes 3D ne doit pas forcément être dense. On privilégie la fiabilité des primitives par rapport à l'exhaustivité de la reconstruction. Les problèmes de sous-détection et de sur-détection de primitives seront résolus dans le processus d'optimisation globale (cf. chapitre 7.3).

On a proposé dans le chapitre 5, une première approche de reconstruction de facettes 3D basée sur les segments 3D. On propose dans ce chapitre une deuxième approche complémentaire, basée sur les régions. Ces deux approches fourniront un ensemble de facettes concurrentes qui sera également utilisé dans le processus d'optimisation globale afin de valider les meilleures hypothèses de facettes 3D en tenant compte de la régularité de la surface finale ainsi que des discontinuités.

6.2 État de l'art en reconstruction de surfaces planes

Dans le contexte satellitaire THR, peu de travaux traitent le thème de la reconstruction de bâti. Dans [FBA01], l'auteur présente des évaluations qualitatives et quantitatives sur des images IKONOS concernant les aspects radiométriques et les précisions de reconstruction de bâti avec des méthodes existantes. Dans [SD01], l'auteur propose une méthode de détection des bâtiments en 2D à partir d'images satellitaires en recherchant les directions dominantes par analyse de Fourier. L'état de l'art portera principalement sur les méthodes utilisées en imagerie aérienne, qu'il faudra adapter au contexte satellitaire.

De nombreux auteurs ont montré l'importance des surfaces planes pour décrire les scènes urbaines [VD01]. Trois différentes approches sont développées dans la littérature.

La première approche est basée sur l'extraction de plans 3D à partir de nuages de points 3D obtenus par focalisation sur un MNE (Modèle numérique d'élévation), soit par une approche bayésienne [CJCP99], soit en utilisant la transformée de Hough 3D pour extraire des hypothèses de plans 3D [JPDP00]. Une étape d'intersection et de fusion des plans extraits permet d'aboutir à un modèle polyédrique du bâtiment. Dans [BZ99], on recherche des demi-plans 3D pivotant autour des segments 3D extraits et la délimitation des facettes 3D est réalisée par intersection des plans 3D avec les segments voisins extraits. Cette approche ne gère pas le problème de sous-détection de primitives et toute la chaîne de traitement dépend de la qualité des segments 3D.

La deuxième approche se base sur une délimitation de l'emprise 2D du bâtiment et la recherche de plans 3D modélisant les toits à l'intérieur de cette emprise. On s'affranchit ainsi du problème de délimitation des facettes 3D et le problème est beaucoup plus contraint. Ces emprises 2D peuvent être des rectangles obtenus à partir des images [NN01] ou du MNE par des processus stochastiques [ODZ03] ou bien en combinant le MNE avec des orthoimages [SF01] [VC02b]. Les rectangles délimitent les bâtiments, la reconstruction 3D est obtenue en modélisant les toits par des modèles plus ou moins complexes et en récupérant les altitudes moyennes à partir du MNE. D'autres approches utilisent des données externes pour fournir les emprises 2D telles que les cadastres. Dans [Jib02], l'auteur utilise l'emprise cadastrale pour détecter la forme des bâtiments. Les plans 3D s'appuyant sur les bords du cadastre sont détectés par une transformée de hough 3D. Une étape d'intersections de plans 3D fournit un ensemble de facettes dans lequel il faut choisir le meilleur sous-ensemble en adéquation avec le MNE et avec une complexité de structure minimale.

La troisième approche est basée sur la segmentation en régions. L'avantage est que l'on a directement la délimitation des facettes. On peut facilement garder l'information topologique. Certaines approches se basent exclusivement sur le MNE. Dans [Ves00], l'auteur propose un algorithme à trois étapes : sélection des modèles de plans 3D sur le MNE (par un algorithme de RANSAC), fusion des hypothèses et enfin sélection des hypothèses sur un critère MDL. On notera que dans ce cas, aucune information topologique n'est reproduite. Dans [AF99], l'auteur utilise également une segmentation en régions à partir du MNE mais il garde l'information topologique en fusionnant les plans adjacents compatibles. Le modèle de bâtiment final est ensuite validé par reprojection et recalage par rapport aux images. Le problème des approches basées sur les MNE issus de la corrélation est surtout le manque de précision des MNE et les délocalisations des bords de bâtiments. Ces inconvénients sont d'autant plus remarquables à partir d'images satellitaires avec un faible rapport B/H . Pour pallier à ce problème d'autres approches combinent MNE et orthoimage de gradient pour marquer les discontinuités à partir des images [TGD03].

Dans notre cas, on cherche plutôt à obtenir une caricature de surface qu'une reconstruction de bâtiments à proprement dit. On ne cherche pas à garder la topologie. Dans la stratégie globale (cf. section 3.3), on propose un processus d'optimisation globale qui inclue les données sources et l'ensemble des primitives 3D extraites pour modéliser les surfaces 3D. On cherche, par conséquent, à détecter des facettes 3D fiables que l'on pourra utiliser dans ce processus. On n'utilisera pas du tout le processus d'intersection et de fusion de plans 3D au cours de notre démarche. On exclut également les approches reposant sur des emprises 2D. Dans un cadre pratique de l'IGN, on pourrait utiliser le cadastre comme donnée externe. Ce problème ne sera pas traité dans cette thèse. On s'oriente par conséquent, vers la troisième approche basée sur la segmentation des images en régions et leur mise en correspondance pour estimer les facettes 3D correspondantes.

6.3 Méthodologie

Dans cet algorithme de reconstruction des facettes 3D, on a choisi une approche par régions basée sur la segmentation des images et la mise en correspondance des régions résultantes. Tout d'abord, les régions constituent des primitives complémentaires aux segments 3D. En effet, les appariements sont plus fiables grâce à la grande taille des régions. Elles permettent de réaliser des estimations robustes de plans 3D et de corriger ainsi l'imprécision altimétrique due au faible rapport B/H . En outre, cette approche permet de reconstruire directement des surfaces tridimensionnelles, ce qui fournit un modèle 3-D beaucoup plus dense que ceux obtenus à partir d'une

analyse en points caractéristiques ou en segments des deux images de la scène. Enfin, la segmentation en régions apporte une information topologique qui permet de propager des contraintes entre régions adjacentes ou aux parties cachées et qui permet de structurer la scène 3D.

Dans notre approche, on a également opté pour une segmentation des images et non du MNE. Compte tenu du contexte satellitaire, le MNE obtenu par des techniques de corrélation [Bai97], présente beaucoup d'inconvénients ; la délocalisation des bords de bâtiments, la précision altimétrique est réduite à cause du faible rapport B/H du couple stéréoscopique. Par conséquent on exclut, dans notre cas, la segmentation du MNE pour l'extraction de plans 3D. On utilisera des segmentations en régions dans les images. Leur appariement sera d'autant plus fiable que les distorsions géométriques sont réduites entre les deux images grâce au faible rapport B/H .

L'approche globale est illustrée dans la figure 6.1. Trois grandes parties se détachent : la segmentation des images, l'appariement des régions et la reconstruction des facettes 3D.

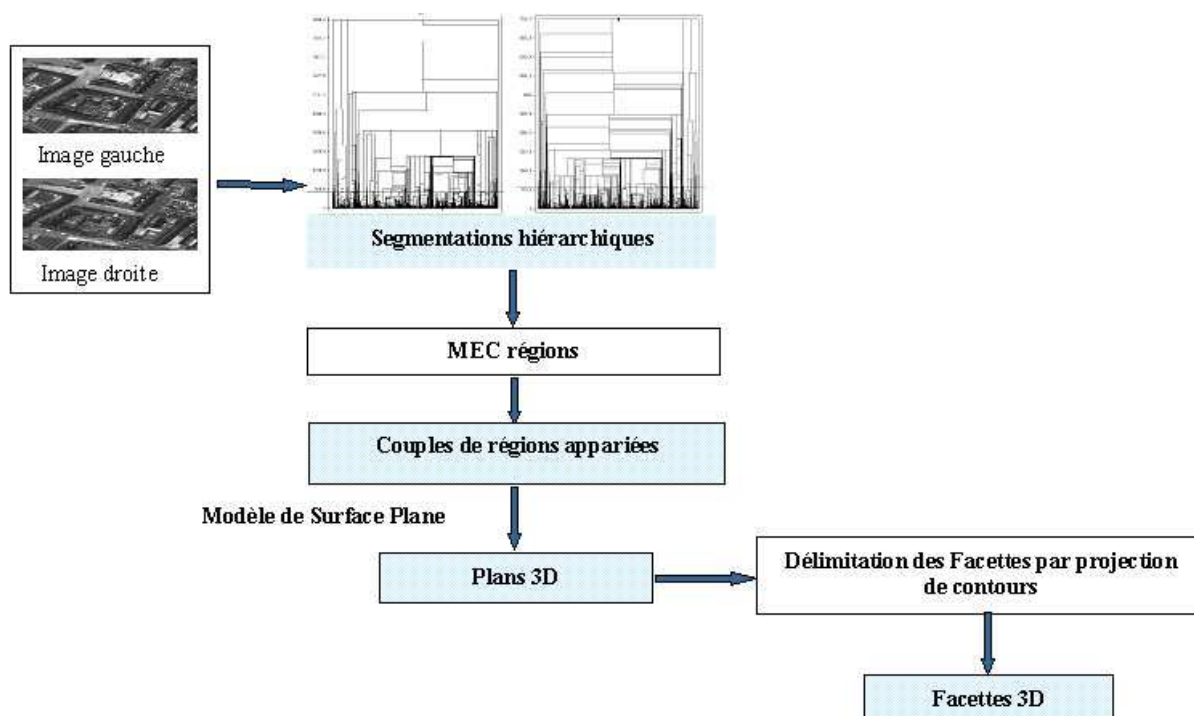


FIG. 6.1 – Schéma global de la reconstruction des facettes 3D par approche région

Tout d'abord, les deux images sont segmentées par des outils de segmentation multi-échelles que l'on décrira dans la section 6.4. La deuxième phase est la mise en correspondance des régions dans les deux images (cf. section 6.5). En supposant un modèle de surface plane, on cherche à reconstruire les plans 3D relatifs aux couples de régions appariées. Les facettes 3D finales sont obtenues par projection des contours des régions sur les plans 3D correspondants.

6.4 Segmentation multi-échelles et hiérarchie

Le but de notre segmentation des images est d'obtenir des partitions en régions denses qui renvoient des primitives 2D interprétables, en vue de les appairer pour définir des facettes. Dis-

posant de deux images différentes d'une même scène, il est très difficile voire impossible d'obtenir deux partitions automatiques dont les régions se correspondent bijectivement : on obtient systématiquement des différences de degrés de découpage à certains endroits de la scène. Pour pallier à ce problème, on propose d'utiliser des segmentations hiérarchiques [GLMC03] qui permettent d'avoir plusieurs niveaux d'interprétation de la scène (pixel, toit, bâtiment...), allant des régions les plus fines aux régions les plus grossières et qui rendent les appariements inter-niveaux possibles. Il suffit dans ce cas, que deux régions qui se correspondent, apparaissent à deux valeurs quelconques du paramètre d'échelle pour qu'elles puissent être appariées. Une approche similaire a été proposée dans [RG92]. En vue d'apparier des images, les auteurs proposent une méthode de division récursive de l'image pilotée par la mise en correspondance des régions, le but étant de produire deux segmentations en régions aussi similaires que possible.

Dans notre approche, les distorsions géométriques et les parties cachées entre les deux images étant réduites, les deux segmentations hiérarchiques sont assez similaires. On se base donc sur deux segmentations initiales des images et on cherchera à les apparier de manière globale. L'appariement de régions est facilité compte tenu du faible rapport B/H . Les images sont tout d'abord segmentées en utilisant un algorithme de segmentation multi-échelles, décrit dans [GLMC03]. Cet algorithme produit des segmentations de l'image $\Omega_0, \Omega_1, \dots, \Omega_n$ par une approche ascendante ou encore "Bottom-Up" (cf. section 6.4.2). En partant d'une sur-segmentation initiale (par l'algorithme de partage des eaux), la hiérarchie de régions est construite par fusions successives des régions initiales suivant le modèle décrit ci-dessous.

6.4.1 Modèle de fusion des régions

Le but est de segmenter l'image en régions correspondant à de véritables facettes 3D donc principalement des toits dans notre cas. Ces régions sont en général de forme simple et de radiométrie homogène. Pour cela, on part d'une formulation classique de la segmentation dans laquelle deux énergies sont mises en compétition : une énergie géométrique et une énergie radiométrique. L'objectif étant de trouver la partition qui minimise l'énergie globale pour un λ donné.

$$E_\lambda(P) = \sum_{R \in P} E_\lambda(R) = \sum_{R \in P} \lambda E_{Géo}(R) + E_{Rad}(R) \quad (6.1)$$

Le λ est un terme de régularisation : pour un fort λ , l'énergie géométrique sera prédominante conduisant à un modèle simple (peu de régions avec des formes simples), alors que si λ est faible on s'attache plus à l'énergie radiométrique et on obtient donc de plus petites régions homogènes vérifiant le modèle radiométrique (cf. Figure 6.4).

En outre, les deux énergies géométrique et radiométrique sont elles-même composées d'une énergie de régularisation (modèle), \mathbf{C} , et une énergie d'attache aux données, \mathbf{D} ,

$$\begin{cases} E_{Géo}(R) = W_G C_{Géo}(R) + D_{Géo}(R) \\ E_{Rad}(R) = W_R C_{Rad}(R) + D_{Rad}(R) \end{cases} \quad (6.2)$$

W_G et W_R constituent des termes de régularisation des deux énergies respectives. Les énergies employées pour la segmentation sont issues du formalisme de codage. Elles comportent une énergie du type codage polygonal pour l'énergie de complexité géométrique et une énergie correspondant au coût d'un codage selon une loi gaussienne i.i.d des valeurs des pixels dans chaque région pour l'aspect radiométrique [Gui03].

6.4.2 Construction de la hiérarchie

Le principe de la segmentation multi-échelles est de rechercher une famille de solutions, pour $\lambda \in \mathbb{R}^+$ qui minimise l'énergie globale E_λ et d'imposer que ces solutions forment une séquence de partitions ordonnées suivant λ , de la plus fine à la plus grossière, modélisant un processus de description multi-échelles.

Au niveau pratique, l'idée est de construire une hiérarchie de régions d'une image à différentes échelles en augmentant à chaque fois l'échelle globale λ et en fusionnant pour cela les régions adjacentes dont la fusion diminue l'énergie globale. Pour la fusion de deux régions adjacentes R_1 et R_2 en une seule région R , on définit λ_{app} l'échelle d'apparition de la région R , qui est la valeur minimale à partir de laquelle on a :

$$E(R_1 \cup R_2) < E(R_1) + E(R_2) \quad (6.3)$$

On obtient ainsi une hiérarchie où chaque nœud est indexé par son échelle d'apparition λ_{app} et correspond à une région définie par la fusion d'un ensemble de régions de base connectées. Une méthode naturelle pour la construction d'une hiérarchie est alors de partir d'une partition fine de l'image I qui constitue les **nœuds de base** de la hiérarchie et de réaliser itérativement la fusion de régions qui a la valeur minimale de l'échelle d'apparition (cf. figure 6.2).

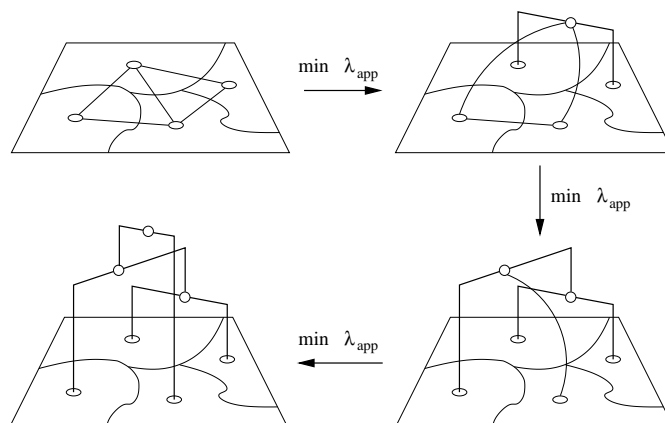


FIG. 6.2 – Construction de la hiérarchie

On définit aussi pour chaque nœud une échelle de disparition λ_{disp} qui correspond en fait à l'échelle d'apparition λ_{app} de la région mère correspondante. Ainsi, à chaque nœud correspond un intervalle $[\lambda_{app}, \lambda_{disp}]$ dans lequel la région R correspondante est visible. Autrement dit, toute coupe de la hiérarchie dans cet intervalle contiendra la région R . Ainsi, une coupe horizontale suivant λ fournit une partition de l'image [GLMC03] (cf. figures 6.3,6.4).

La hiérarchie H peut être illustrée comme un arbre où l'axe vertical représente les échelles d'apparition λ_{app} . Les nœuds de base sont obtenus par sur-segmentation de l'image, la racine (i.e le sommet) correspond à l'image entière. La figure 6.4 illustre différentes coupes de la hiérarchie à des λ croissants.

Les définitions relatives à une hiérarchie sont rappelées dans l'annexe C.1.3.

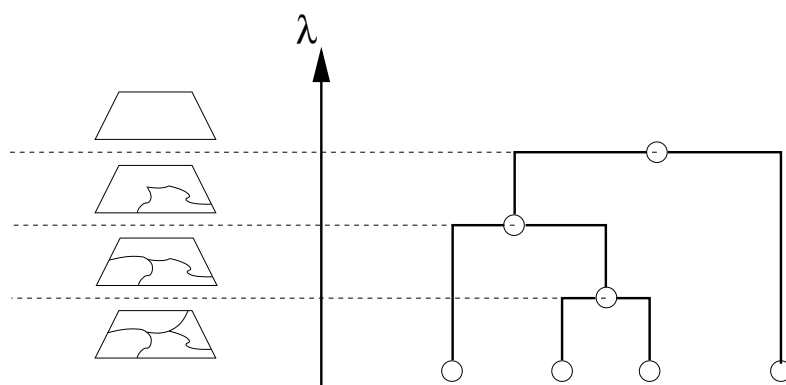


FIG. 6.3 – Différentes échelles de segmentation de la hiérarchie

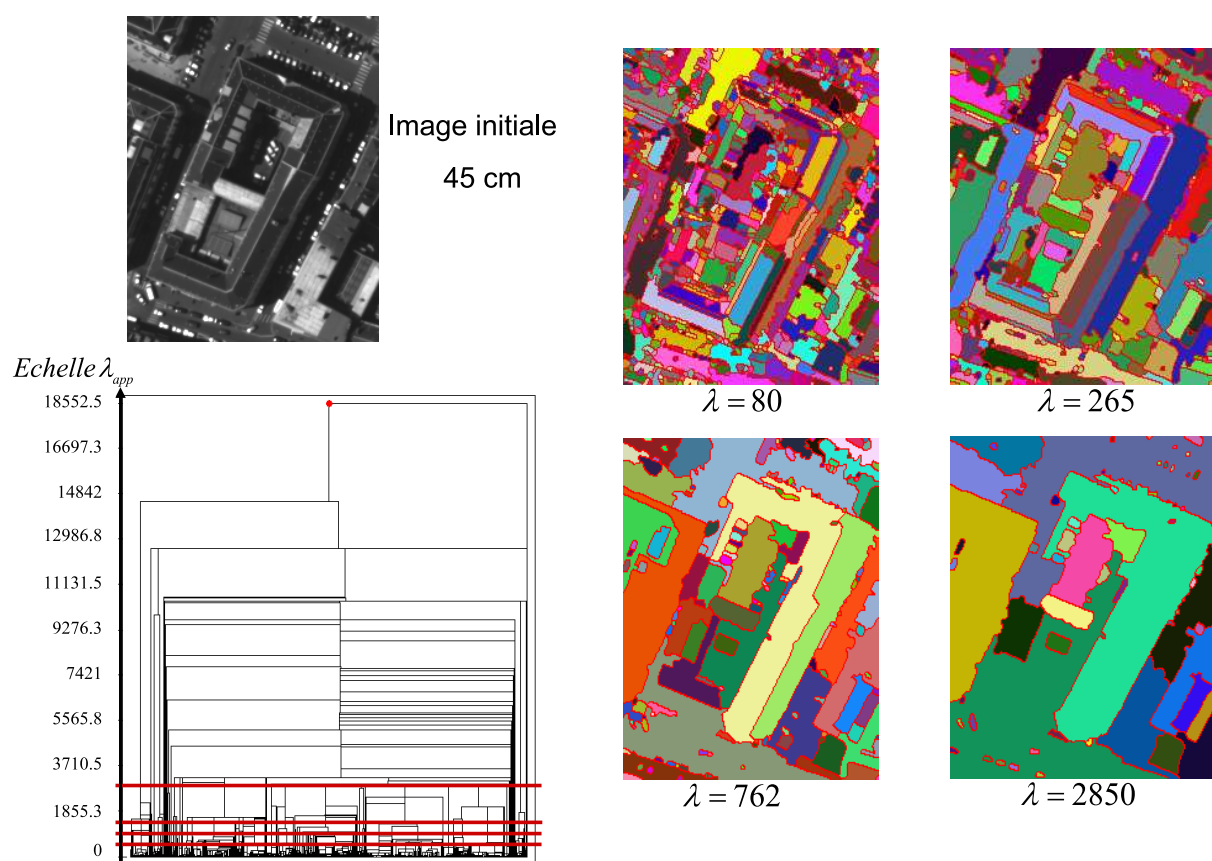


FIG. 6.4 – Segmentation multi-échelles

6.5 Mise en correspondance des régions

La segmentation hiérarchique décrite dans la section précédente fournit des descriptions multi-échelles des deux images. La deuxième étape est la mise en correspondance des régions entre les deux hiérarchies. Dans cette section, un état de l'art des stratégies de mise en correspondance des primitives sera présenté et la méthodologie adoptée sera également détaillée.

6.5.1 État de l'art de la mise en correspondance de primitives

6.5.1.1 Les stratégies d'appariement

Dans la littérature, plusieurs stratégies d'appariement de primitives ont été développées dont les principales sont décrites ci-dessous. [Bai97] propose un état de l'art détaillé de ces stratégies. Les algorithmes cités traitent différents types de primitives : coins, segments, régions, etc.

- ▶ **La programmation dynamique :** Cette méthode a été présentée dans le cas des appariements de pixels (cf. section 2.3.1.2). Elle peut être utilisée pour la mise en correspondance de primitives [OK85, SB94, IB95, Bai97].
- ▶ **L'algorithme de relaxation :** C'est un algorithme de recherche itératif. Le problème de mise en correspondance est traité comme un problème d'étiquetage où chaque primitive de l'image gauche doit avoir au plus une primitive homologue dans la seconde image vérifiant certaines contraintes. La vraisemblance de chaque appariement est représentée par une probabilité. À l'initialisation, les probabilités sont choisies en fonction des critères de ressemblance. L'algorithme va faire évoluer ces probabilités (que l'on associe aux noeuds d'un graphe) de manière itérative en fonction des appariements voisins et des contraintes de cohérence avec le voisinage. Ce sont principalement les règles de mise à jour des probabilités qui changent entre les différents algorithmes de relaxation proposés.
- ▶ **Recherche dans un graphe :** Cette approche nécessite une description symbolique de chacune des images sous la forme d'un graphe d'adjacence, dont les noeuds sont les primitives à appairer [HS89]. Un graphe de correspondance exhaustif est alors créé : les appariements potentiels entre primitives forment les noeuds, les compatibilités entre hypothèses constituent les arcs. La recherche d'une mise en correspondance cohérente équivaut à trouver la plus grande composante connexe du graphe de correspondance.
- ▶ **Prédiction et vérification d'hypothèses :** C'est une technique récursive pour appairer les primitives, au cours de laquelle les appariements valides sont utilisés pour générer des hypothèses d'appariement (propagation basée sur les relations de voisinage entre primitives), les conflits sont résolus par sélection des hypothèses les plus fiables. Cette stratégie est appliquée pour l'appariement des chaînes [HF92], des segments [AF87], ou des régions [CG87].

Dans notre cas, on dispose d'une segmentation multi-échelles des images qui fournit plusieurs niveaux d'interprétation. Ces segmentations ont l'avantage de présenter une architecture spécifique des graphes sous forme de partitions emboîtées. La recherche exhaustive dans un graphe de correspondance entre les deux hiérarchies augmenterait considérablement les temps de calcul puisque les segmentations initiales sont déjà sous forme multi-échelles. Pour mettre à profit l'architecture emboîtée de la hiérarchie, on opte pour une stratégie de prédiction et de vérification d'hypothèses. Les hypothèses d'appariement ne seront pas propagées par des relations de voisinage mais par un processus multi-échelles, ou encore "fils-mère" qui consiste à propager les

contraintes d'appariement de bas en haut de la pyramide suivant la propriété suivante : une région mère vérifie une contrainte si tous ses fils vérifient la même contrainte. La propagation ascendante des contraintes rend les algorithmes beaucoup plus robustes aux erreurs de segmentations et au bruit comme on le verra dans la section 6.5.2.3.

6.5.1.2 Les stratégies de contrôle et de validation

Parallèlement aux stratégies d'appariement, il existe des stratégies de contrôle pour guider ou valider l'appariement :

- ▶ **Stratégie multi-échelles** : Elle repose sur le fait que la disparité évolue linéairement en fonction de l'échelle, et consiste à utiliser la disparité obtenue à un niveau de résolution donné pour guider l'appariement au niveau supérieur. La réduction de l'échelle permet la réduction de l'espace de recherche, d'où un appariement plus facile (mais moins précis). Les stratégies multi-résolutions en sont un cas particulier ; elles consistent à sous échantillonner les images à des résolutions différentes [Han89].
Ces stratégies de contrôle améliorent l'efficacité des algorithmes et réduisent le risque de faux appariements. Elles sont cependant sensibles aux propagations d'erreurs d'une échelle à l'autre, notamment en cas de fortes discontinuités.
- ▶ **Stratégie hiérarchique** : Il s'agit d'apparier des primitives dont les niveaux de représentation sémantique sont différents ; l'appariement des structures haut niveau, peu ambigu, guide l'appariement des primitives inférieures [MT94]. Contrairement aux approches multi-échelles, la réduction de densité des primitives est obtenue par une abstraction de structures haut-niveau plutôt que par une réduction de l'échelle, et prend ainsi en compte la cohérence structurelle de la scène.
- ▶ **Appariement bijectif** : Les deux images du couple ayant souvent un rôle dissymétrique dans l'appariement, une stratégie de validation consiste à appliquer l'algorithme en échangeant le rôle des images, et à ne valider que les appariements communs aux deux expériences [Han89].

Dans notre cas, pour guider les appariements, on a opté pour une stratégie hiérarchique adaptée à la segmentation multi-échelles. Les appariements seront réalisés par approche descendante "Top-Down" en commençant par le niveau le plus grossier de la hiérarchie et seront validés par des critères relatifs entre régions mère-fils. Les deux hiérarchies sont traitées de manière symétrique. L'ensemble d'associations final sera validé par un critère d'unicité.

La mise en correspondance des régions a pour but d'apparier globalement les deux hiérarchies et est composée de trois étapes (cf. figure 6.5).

La première étape consiste à apparier les régions de chaque hiérarchie avec la seconde hiérarchie. On disposera ainsi, pour chaque hiérarchie, de correspondances 1 à N , le but étant de toujours avoir le bon homologue parmi la liste d'homologues possibles. La deuxième étape sera la qualification des appariements en affectant une qualité à chaque couple de régions. L'objectif final sera de trouver les meilleurs couples appariés dans les deux hiérarchies, qui renvoient des régions correspondant à de véritables facettes 3D. Les couples seront réalisés en combinant les qualités d'appariement et la contrainte de planéité des régions. Comme le montre la figure 6.5, trois étapes se détachent pour la mise en correspondance des régions : l'appariement des régions, la qualification des appariements et la recherche des couples dans la hiérarchie pour la validation des appariements. Ces étapes seront détaillées dans les sections suivantes.

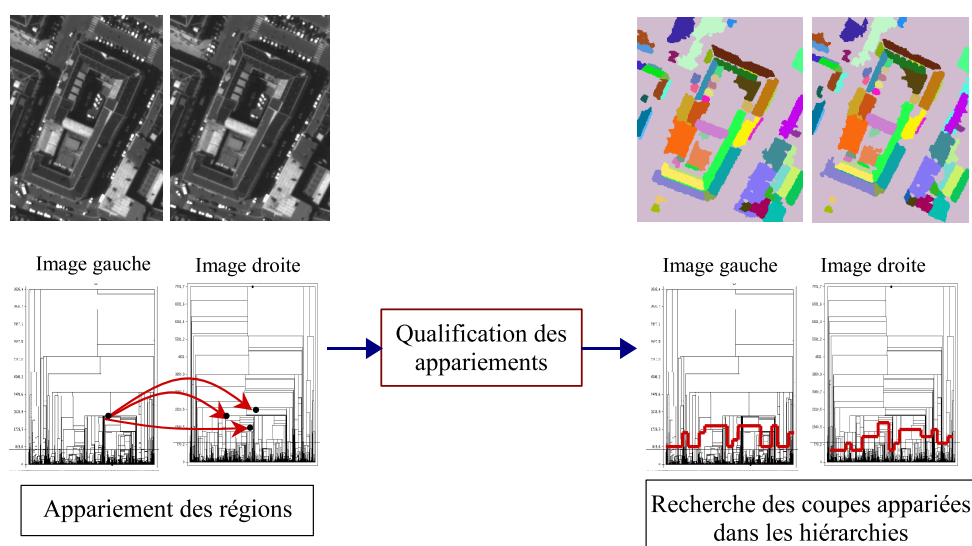


FIG. 6.5 – Mise en correspondance des régions et détection des facettes 2D

6.5.2 Appariement d'une région

Dans un premier temps, les régions d'une hiérarchie sont appariées indépendamment les unes des autres avec les régions de la seconde hiérarchie. Le processus est réalisé sur les deux hiérarchies de façon symétrique. L'appariement de régions est un problème classique en vision [SVC89, RG92]. Pour cette étape d'appariement, on utilise plusieurs contraintes pour réduire le nombre d'homologues à traiter ; contrainte de taille, contrainte épipolaire et altimétrique, contraintes de similarité et de recouvrement. L'originalité de notre approche réside dans la propagation ascendante des contraintes d'appariement dans les hiérarchies. Ceci rend la recherche d'homologues plus robuste.

Dans la suite de l'étude, on adoptera les notations suivantes : Ω_g et Ω_d désignent l'ensemble des régions de la hiérarchie de l'image gauche et de l'image droite.

6.5.2.1 Contrainte de taille

Cette contrainte est utilisée tout au long du processus pour éviter de traiter des régions jugées trop petites ou trop grandes. En fonction du niveau d'interprétation souhaité (toits, cheminées, chiens assis...) et de la résolution de l'image, on impose un seuil minimal et maximal pour la taille des régions, S_{min}^t et S_{max}^t qui permet d'élaguer des nœuds de la hiérarchie qui ne sont pas susceptibles de représenter des facettes d'intérêt.

6.5.2.2 Contrainte épipolaire et altimétrique

Pour rechercher la liste d'homologues possibles pour une région donnée, on commence par appliquer la contrainte épipolaire sur le barycentre G de la région (cf. figure 6.6(a)). Il faut que le barycentre de la région homologue appartienne à un bandeau épipolaire obtenu avec une tolérance préfixée par rapport à la ligne épipolaire correspondante à G qu'on notera L_G . L'ensemble des régions homologues, d'une région donnée R , par la contrainte épipolaire sera noté $HEpi(R)$.

Afin de réduire l'espace de recherche, on utilise également une contrainte altimétrique. On dispose en entrée d'un Modèle Numérique d'Élévation en $2D\frac{1}{2}$ calculé par des techniques de

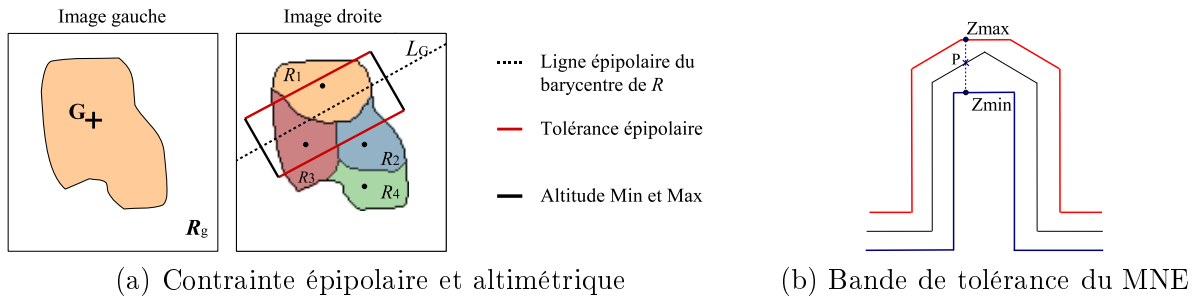


FIG. 6.6 – Contrainte épipolaire et altimétrique

corrélation [Bai97]. Le MNE obtenu n'est pas précis à cause du faible B/H et de la taille des fenêtres de corrélation mais il a l'avantage d'être robuste. Grâce à un volume dilaté du MNE, on dispose d'une bande de tolérance qui renvoie l'altitude Min et Max de chaque pixel (cf. figure 6.6(b)).

La recherche des homologues épipolaires d'une région revient ainsi à retenir tous les nœuds de la seconde hiérarchie qui sont contenus dans le rectangle épipolaire correspondant.

Cependant, la contrainte épipolaire ¹ ainsi appliquée présente deux principaux inconvénients :

► **Risque de perte de régions homologues**

- Si les régions à appairer ne sont pas segmentées de la même façon dans les images (cf. figure 6.6(a)).
- Dans le cas d'un barycentre excentré, le paramètre de tolérance épipolaire devient critique. Des régions homologues peuvent être perdues.

► **La contrainte n'est pas robuste au bruit.** Des régions parasites peuvent se rajouter au cours des fusions successives et apparaître parmi les homologues possibles.

Afin de faibliser les appariements, notre contribution majeure consiste à propager la contrainte épipolaire de manière ascendante dans la hiérarchie ce qui rend la mise en correspondance beaucoup plus robuste comme on le verra dans la section suivante.

6.5.2.3 Propagation ascendante de la contrainte épipolaire d'une région

Soit une région de référence N_{Ref} , $\mathcal{B}(N_{Ref})$ l'ensemble de ses nœuds de base. L'étape d'initialisation consiste à marquer les homologues de tous ses nœuds de base dans la seconde hiérarchie. L'ensemble des homologues $HEpi(N_{Ref})$ sera obtenu par propagation ascendante de la contrainte épipolaire, à partir des nœuds marqués dans la seconde hiérarchie. Le principe de la propagation de la contrainte étant le suivant :

Un nœud père est considéré comme homologue d'un nœud de référence par la contrainte épipolaire si tous ses fils vérifient cette propriété.

L'algorithme 1 détaille l'application de la contrainte épipolaire à une seule région de référence. La figure 6.7 schématise la propagation de la contrainte épipolaire pour le nœud N_{Ref} . L'ensemble des homologues $HEpi(N_{Ref})$ est encerclé dans la seconde hiérarchie.

Rappels :

- Les fils d'un nœud x sont notés $\mathcal{F}(x)$. Ses nœuds de base sont notés $\mathcal{B}(x)$.

¹Par abus de langage, la contrainte épipolaire sous-entend la contrainte épipolaire et altimétrique qui permet d'avoir le rectangle épipolaire homologue.

Algorithme 1: Propagation ascendante de la contrainte épipolaire : Cas d'une région**Entrée :** Deux hiérarchies H_1 et H_2 , $X \in H_1$ **Sortie :** $HEpi(X) \subset H_2$: Les homologues de X par la contrainte épipolaire**Initialisation :****pour** $x \in \mathcal{B}(X)$ **faire** Trouver dans H_2 les homologues de x par contrainte épipolaire : $HEpi(x)$ **pour chaque** $y \in HEpi(x)$ **faire** Marquage de y $front \leftarrow y$ $HEpi(X) \leftarrow y$ **Propagation ascendante dans H_2 :**1 **tant que** $front$ non vide **faire** $front \rightarrow x$ $p(x)$ père du nœud x **si** $p(x)$ a déjà été traité **alors**

go to 1

sinon **pour chaque** $y \in \mathcal{F}(p(x))$ **faire** **si** y n'est pas marqué **alors**

go to 1

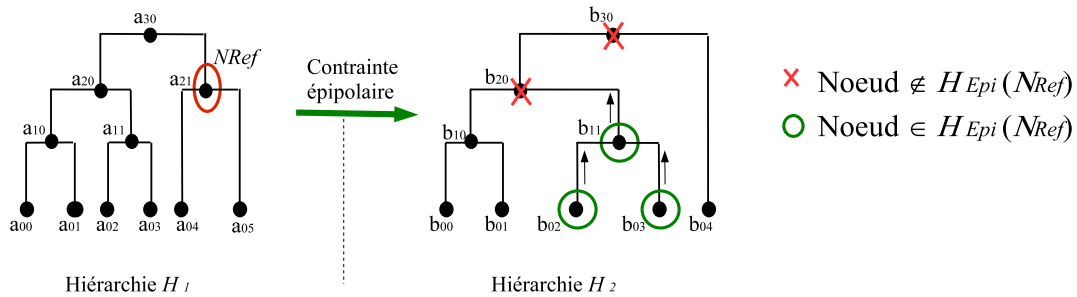
 Marquage de $p(x)$ $front \leftarrow p(x)$ $HEpi(X) \leftarrow p(x)$ 

FIG. 6.7 – Application de la contrainte épipolaire par propagation ascendante

La propagation ascendante de la contrainte épipolaire consitute une contribution majeure dans notre travail. La mise en correspondance des régions est plus robuste. Elle évite de perdre des régions homologues, par exemple dans le cas de barycentre excentré. De plus, le seuil de tolérance à la ligne épipolaire n'est plus un paramètre critique puisqu'il est juste utilisé pour les nœuds de base. Enfin, elle évite de considérer comme homologues des régions parasites, dues au bruit, qui se rajoutent au cours des fusions de régions (exemple figure 6.7, noeud b_{04}).

6.5.2.3.1 Extensions : En pratique, la contrainte épipolaire est appliquée à toutes les régions d'une hiérarchie de référence. Afin d'optimiser le temps de calcul, le traitement de toutes les régions de référence se fera par propagation ascendante dans la hiérarchie. L'algorithme est détaillé dans l'annexe A.2.

6.5.2.4 Contrainte de similarité

A ce niveau, chaque région R a une liste d'homologues $HEpi(R)$ dans la seconde hiérarchie. Pour réduire la combinatoire, on utilise une contrainte de similarité [SVCG89]. L'ensemble des régions homologues de R par la contrainte de similarité sera noté $HSim(R)$.

$$Sim(R_g, R_d) = \sum_{p=1}^q w_p s_p(R_g, R_d) \quad (6.4)$$

s_p est le score de similarité pour chaque paramètre ou attribut de la région. w_p est le poids affecté au score de similarité s_p défini par l'équation 6.5. Si les attributs des régions sont similaires, le score tend vers 0.

$$s_p(R_g, R_d) = 1 - \frac{\min(A_p(R_g), A_p(R_d))}{\max(A_p(R_g), A_p(R_d))} \quad (6.5)$$

A_p étant un attribut de la région. Dans notre cas, on utilise trois attributs qui sont la surface, les dimensions du rectangle englobant et le moment spatial d'ordre 2. Ces attributs de forme suffisent pour avoir un premier ensemble d'associations. Les caractéristiques radiométriques des régions seront utilisées pour la qualification des appariements. Le score de similarité doit être supérieur à un seuil préfixé S^s . Le seuil de similarité est un paramètre critique, difficile à évaluer puisqu'il dépend de la ressemblance des segmentations des deux images. En effet, si les segmentations sont très semblables, les scores de similarité tendent vers 0. En pratique, on choisit un seuil tolérant $S^s = 0.3$ pour éviter de perdre des régions homologues.

6.5.2.5 Contrainte de recouvrement

Le principe de la contrainte de recouvrement est illustré figure 6.8. La reconstruction du plan 3D sera détaillée dans la section 6.5.3.1. En projetant les deux régions sur le plan $\Pi_{(R_g, R_d)}$, on obtient deux régions rectifiées R_g^π et R_d^π . Le recouvrement est égal au rapport de l'intersection des deux régions rectifiées sur leur union. Pour valider l'appariement, il faut que le recouvrement des deux régions pojetées soit supérieur à un seuil S^r . Ce seuil est plus facile à interpréter que le seuil de similarité puisqu'il ne dépend pas des segmentations. On choisit un seuil strict $S^r = 60\%$.

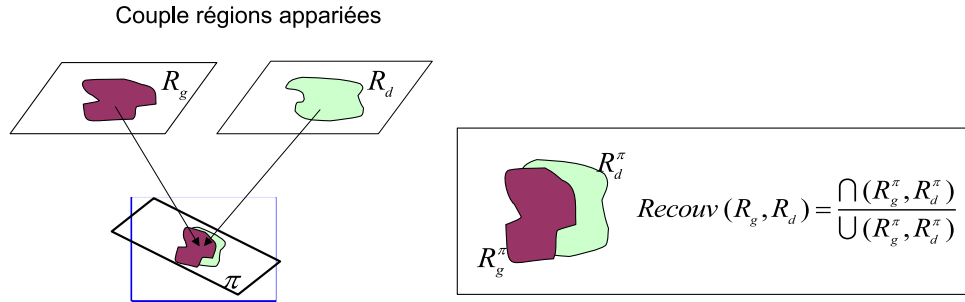


FIG. 6.8 – Contrainte de recouvrement

6.5.2.6 Formulation de la relation d'appariement

On définit alors la relation d'appariement \mathcal{H} entre deux régions comme suit :

$$\forall R_g \in \Omega_g, \forall R_d \in \Omega_d \quad R_g \mathcal{H} R_d \iff$$

$$\begin{cases} 1. S_{min}^t \leq R_g \leq S_{max}^t \text{ et } S_{min}^t \leq R_d \leq S_{max}^t \\ 2. R_d \in HEpi(R_g) \\ 3. R_d \in HSim(R_g) \iff Sim(R_g, R_d) \geq S^s \\ 4. R_d \in HRecouv(R_g) \iff Recouv(R_g, R_d) \geq S^r \end{cases} \quad (6.6)$$

Les régions homologues doivent vérifier de manière séquentielle la contrainte de taille, la contrainte épipolaire, la contrainte de similarité et la contrainte de recouvrement.

6.5.2.7 Extensions :

- **Contrainte d'adjacence** : D'autres contraintes peuvent être appliquées pour tirer profit de la topologie fournie implicitement par la segmentation dense des images telles que la contrainte d'adjacence qui permet de propager l'information d'appariement aux régions voisines.

Soit une région R non appariée, $Adj(R)$ l'ensemble de ses régions adjacentes et l'ensemble de leurs homologues respectifs par la contrainte épipolaire. Une relation d'adjacence partielle entre les régions correspondantes dans la seconde hiérarchie renvoie les homologues de R par la contrainte d'adjacence : $HAdj(R)$ qui est l'ensemble des régions qui satisfont la contrainte épipolaire et qui ont au moins une adjacence avec les homologues de $Adj(R)$ (cf. figure 6.9). L'utilisation de cette contrainte permettrait de maximiser le nombre d'appariements au niveau d'une hiérarchie. La relation d'adjacence est choisie partielle, afin de traiter le cas des parties cachées d'une image à l'autre.

- **Focalisation sur les bâtiments** : Afin de réduire le temps de calcul, une étape de focalisation sur les bâtiments est nécessaire. On propose d'appliquer un masque de sol au niveau de la hiérarchie. Ce masque peut être obtenu par des techniques automatiques de classification sol/sur-sol [Bai97] ou bien par des méthodes semi-automatiques, telles que la croissance de régions à partir d'un germe choisi manuellement ou encore l'utilisation de données externes telles que les limites cadastrales ou le réseau routier 3D pour estimer l'altitude du sol. Ainsi, seules les régions appartenant aux toits seront traitées.

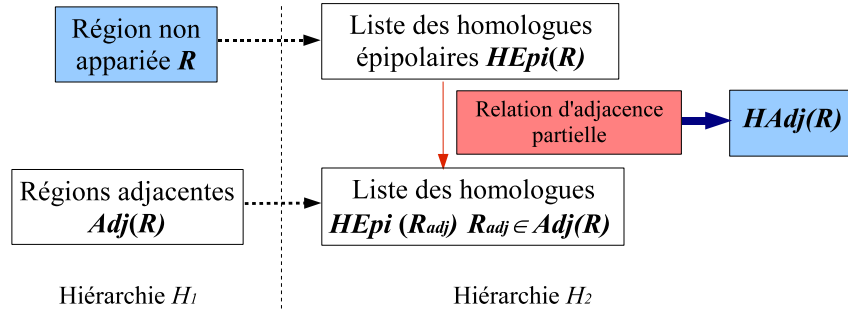


FIG. 6.9 – Contrainte d'adjacence inter-pyramides pour les régions non appariées

6.5.3 Reconstruction des plans 3D

La reconstruction du plan 3D d'un couple de régions appariées est indispensable pour calculer le recouvrement des régions (cf. section 6.5.2.5), pour qualifier la qualité de l'appariement (cf. section 6.5.4.2), et bien sûr pour la reconstruction finale de la scène 3D. La reconstruction du plan 3D peut être photométrique, en utilisant des attributs radiométriques des régions ou bien géométrique [VT95].

Pour notre problème, on a testé deux approches pour reconstruire les plans 3D : soit à partir de la mise en correspondance des contours des deux régions, soit par une estimation robuste à partir du MNE initial. Ces approches seront détaillées dans les sections suivantes.

6.5.3.1 Estimation du plan 3D par mise en correspondance des contours des régions

La mise en correspondance des contours de régions est réalisée dans l'espace des disparités. Les contours sont rectifiés en géométrie épipolaire de telle façon que deux points de contour homologues (u_g, v_g) et (u_d, v_d) se trouvent sur la même ligne épipolaire donc $v_g = v_d = v$. Les contours sont alors appariés en respectant la contrainte d'ordre. L'appariement des points se fait par corrélation. Dans le contexte satellitaire, pour des couples à faible B/H , il est indispensable d'avoir une précision subpixelaire au niveau de l'appariement pour améliorer la précision altimétrique de reconstruction. Pour cela, l'image cible est rééchantillonnée. Les valeurs de pixels dans une vignette corrélation sont recalculées par un interpolateur bicubique. On obtient alors une courbe de corrélation pour un pixel homologue, le pic maximum renvoie une position subpixelaire du point.

L'estimation des plans 3D est réalisée, comme la mise en correspondance des contours, dans l'espace des disparités. Dans l'espace rectifié, l'équation du plan support de la face est décrit sous la forme suivante :

$$au_g + bu_d + cv + d = 0 \quad (6.7)$$

Par conséquent, un point de l'image gauche (u_g, v) est visible aussi dans l'image droite aux coordonnées $(-\frac{au_g + cv + d}{b}, v)$. La relation est linéaire et la dépendance avec les caractéristiques du plan est simple :

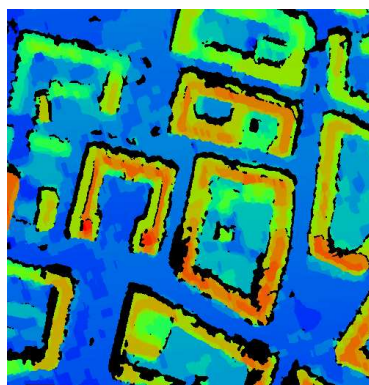
$$\begin{pmatrix} u_d \\ v \end{pmatrix} = \begin{pmatrix} \frac{a}{b} & \frac{c}{b} \\ 0 & 1 \end{pmatrix} * \begin{pmatrix} u_g \\ v \end{pmatrix} + \begin{pmatrix} \frac{d}{b} \\ 0 \end{pmatrix} \quad (6.8)$$

Le calcul des paramètres du plan se fait par des M-estimateurs L_p (cf. Annexe B.3.2). En pratique, $p = 1.2$ représente un bon compromis entre complexité et stabilité [ZF92]. Bien que moins robustes que d'autres estimateurs (RANSAC, moindres carrés médians), les M-estimateurs donnent de bons résultats avec une bonne initialisation. Dans notre cas, l'initialisation dépend des régions appariées, donc le comportement des M-estimateurs permettra de qualifier le processus d'appariement. L'inconvénient de cette approche est que l'équation du plan 3D dépend de la similarité des segmentations des images.

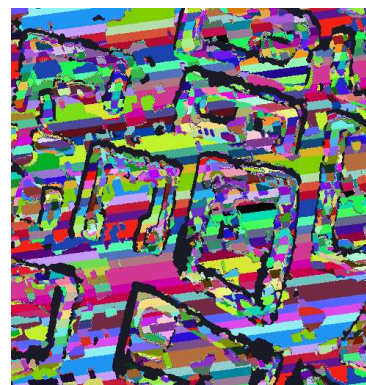
On notera également que l'équation 6.8, peut être utilisée pour réaliser une reconstruction photométrique des facettes en injectant dans l'équation des attributs des deux régions et non les points de contours appariés [VT95].

6.5.3.2 Estimation du plan 3D à partir du MNE

La deuxième approche pour estimer les plans 3D est l'estimation robuste des paramètres de plans à partir du MNE [Bai97]. En effet, on dispose en entrée d'un MNE non précis, mais robuste. La segmentation en régions de ce MNE est difficile à réaliser en utilisant les altitudes. En effet, le MNE est calculé par programmation dynamique, l'inconvénient de cette méthode est la difficulté de propager l'information inter-lignes épipolaires. La carte de disparité finale est calculée par bandes le long des épipolaires (cf. figure 6.10).



(a) MNE Baillard



(b) MNE labellisé

FIG. 6.10 – Inconvénients du MNE de corrélation

Pour cette raison, on a opté pour une segmentation en régions dans l'image. La délimitation des régions est obtenue à partir des images. Le plan 3D correspondant est alors estimé de manière robuste à partir du MNE. Une estimation robuste par les M-estimateurs $L_{1,2}$ (cf. Annexe B.3.2) permet de calculer l'équation du plan sous la forme $z = ax + by + d$.

Cette deuxième approche a été retenue puisque l'estimation est plus robuste que dans le cas de la mise en correspondance des contours de régions. L'estimation du plan est réalisée à partir d'un grand nombre de points communs aux deux régions et pas uniquement à partir des points de contours. Cette approche est également moins sensible aux problèmes de segmentations des régions.

6.5.4 Qualification des appariements

La deuxième étape dans la mise en correspondance des deux hiérarchies est la qualification des différents appariements obtenus. Contrairement à la segmentation, l'appariement est réalisé par une approche descendante. L'appariement est en effet plus fiable sur des grandes régions, on commence donc par le niveau le plus grossier de la hiérarchie. Chaque appariement est qualifié par le score de corrélation des deux régions appariées. On utilise pour cela un volume de corrélation dans l'espace objet. Le calcul du volume de corrélation et la qualification des appariements sont détaillés dans les prochaines sections.

6.5.4.1 Volume de corrélation

Le volume de corrélation (cf. figure 6.11) est calculé dans l'espace objet, en coordonnées géographiques [PTJ00]. L'emprise planimétrique de ce volume correspond à la zone de recouvrement du couple stéréoscopique. On estime ensuite un intervalle d'altitude en utilisant un modèle numérique d'élévation calculé sur le couple. Une fois le volume discrétisé, on attribue à chacun de ses voxels un score d'attache aux données quantifiant l'appartenance de ce voxel à la scène 3D que l'on cherche à reconstruire. Pour cela, on reprojette le voxel et ses voisins horizontaux sur une fenêtre de taille fixée, et on mesure la ressemblance des imagerie par le coefficient de corrélation normalisé centré. On a choisi comme taille de fenêtre 3×3 .

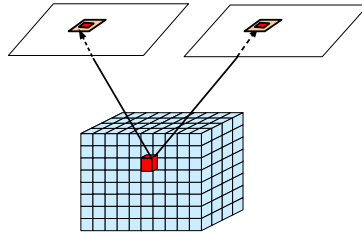


FIG. 6.11 – Calcul du cube de corrélation

Discrétisation du cube : Pour discrétiser le cube en planimétrie, on utilise la résolution de l'image $\Delta x = \Delta y = r_{img}$. La discrétisation en z est plus délicate. En effet, une erreur dans la valeur z d'un pixel (x, y) dans le cube, entraîne des déplacements $\delta x, \delta y$ lors de la reprojection dans les images et par conséquent les scores de corrélation seraient dégradés.

En rectifiant les images en géométrie épipolaire, on a la relation suivante entre la résolution de l'image r_{img} et le pas en z (Δz) : $\Delta z = r_{img} * \frac{B}{f}$ (cf. Eq 2.2). En supposant un déplacement horizontal maximal d'un pixel suivant les lignes épipolaires $r_e = r_{img}$, et que la direction épipolaire fasse un angle θ avec l'axe des x , les déplacements engendrés suivant x, y seront $\delta x = r_e * \cos(\theta)$ et $\delta y = r_e * \sin(\theta)$. Ces déplacements subpixellaires certes, dégradent tout de même les scores de corrélation entre points homologues. Il faut donc discrétiser plus finement en z . En pratique on choisit $\Delta z = \Delta xy = r_{img}$ pour deux raisons principales :

- Ce pas de discrétisation correspond à un rééchantillonnage subpixellaire au $\frac{B}{H}$ ème près. Dans notre contexte, le rapport $\frac{B}{H}$ varie entre 0.05 et 0.2. Ce qui revient à dire que le cube est discrétisé respectivement au $\frac{1}{20}$ ème et $\frac{1}{5}$ ème de pixel en z . Ce cube sera utilisé pour la qualification des facettes 3D (cf. section 6.5.4.2).
- Le fait de choisir le même pas de discrétisation en xy et en z permet d'avoir un cube homogène dans l'espace objet. Ceci est important pour la construction du graphe 3D à partir du cube où l'on aura la même distance entre les différents noeuds (cf. chapitre 7.3).

6.5.4.2 Qualification de l'appariement de deux régions

Un couple de régions appariées est qualifié par le score de corrélation moyen de la facette 3D correspondante \mathcal{F} . La délimitation de la facette 3D est obtenue par intersection des deux régions rectifiées sur le plan correspondant. La facette est validée si et seulement si elle appartient à la bande de tolérance autour du MNE (cf. section 4.3.3).

L'intersection de la facette 3D et du cube précalculé, renvoie les scores de corrélation sur tous les voxels de la facette (cf. figure 6.12).

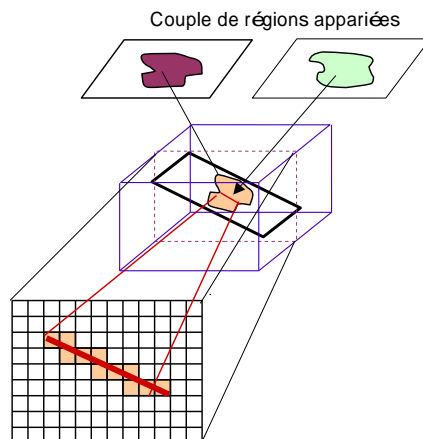


FIG. 6.12 – Calcul du score d'une facette 3D

L'appariement est qualifié par la moyenne des scores de corrélation sur la facette.

$$Q(R_g, R_d) = \frac{\sum_{\mathcal{V} \in \mathcal{F}} \text{ScoreCorrel}(\mathcal{V})}{\text{Card}(\mathcal{F}_{(R_g, R_d)})} \quad (6.9)$$

$\mathcal{F}_{(R_g, R_d)}$ étant discrétisée et considérée comme un ensemble de voxels \mathcal{V} .

Le cube de corrélation permet un gain de temps important lors de la qualification des appariements puisque les scores sont calculés une seule fois sur tous les voxels du cube. Cette méthode est adaptée aux faibles B/H puisqu'il n'y a pas beaucoup de distorsions entre les images. En effet, dans notre contexte, les vignettes horizontales suffisent pour qualifier les appariements. On n'a pas besoin de rectifier les régions pour les corrélérer dans l'espace objet d'où un gain de temps important.

6.5.5 Appariement global des hiérarchies

L'appariement global des hiérarchies consiste à obtenir deux coupes dans les deux hiérarchies qui permettent d'avoir la meilleure segmentation de la scène en facettes planes et qui soient appariées entre elles.

Une coupe dans une hiérarchie H est un ensemble de nœuds que toute branche de H intersecte une seule fois. Elle constitue une partition dense de l'image en régions. Or, en utilisant un masque sur le sol et les qualités d'appariement des régions comme critère de validation, l'ensemble de régions retenues dans une hiérarchie ne constituera pas une coupe, à cause des parties cachées, des régions non appariées et des nœuds appartenant au sol qui ne seront pas traités. Notre problème revient alors à rechercher deux segmentations optimales non denses dans les deux hiérarchies, appariées entre elles et qui renvoient l'ensemble des facettes 3D planes.

La recherche de la segmentation optimale est détaillée dans ce qui suit pour une hiérarchie. En pratique, le processus est appliqué aux deux hiérarchies afin de traiter le couple de façon symétrique et d'obtenir des appariements plus fiables (cf. section 6.5.5.2).

6.5.5.1 Recherche de la segmentation optimale dans une hiérarchie

Les caractéristiques de la segmentation optimale Ω_{op} dans une hiérarchie sont les suivantes :

$$\begin{cases} \forall i, j, i \neq j \ R_i \cap R_j = \emptyset \\ \Omega_{op} = \cup R_i \subset Image \end{cases} \quad (6.10)$$

Contrairement à la segmentation et à l'appariement des régions, la validation des appariements ou encore la recherche de la segmentation optimale se fait par une approche descendante ou "Coarse To Fine". Les appariements sont en effet plus fiables sur des grandes régions, on commence donc par le niveau le plus grossier de la hiérarchie. Cette stratégie de validation combine des critères photométriques et un critère de planéité des régions. Ces critères n'ayant pas été utilisés pour la phase d'appariement.

On rappelle ici quelques notations qui seront utilisées dans la hiérarchie :

- $H(R)$ est l'ensemble des nœuds homologues de la région R .
 - $Fils(R)$ désigne l'ensemble des nœuds fils de la région R .
- $$\forall R \in H, Fils(R) = \{R_i \in H / \cup R_i = R\}.$$

On définit pour chaque région R , une qualité $Q(R)$ qui qualifie l'ensemble de ses appariements et qui correspond au maximum des scores de corrélation renvoyés avec l'ensemble de ses régions homologues.

$$Q(R) = \max(Q(R, R'_i)) \ R'_i \in H(R) ; i = 0, \dots, Card(H(R)) \quad (6.11)$$

Pour une description multi-échelles, la formule 6.10 implique :

$$\text{Si } R \in \Omega_{op}, \forall R_f \in Fils(R), R_f \notin \Omega_{op} \quad (6.12)$$

La recherche de la segmentation optimale se fait par une approche descendante. Les régions retenues pourront appartenir à différents niveaux de la hiérarchie. D'après l'équation 6.12, on retient une région à un niveau donné si les deux conditions suivantes sont réalisées :

Condition 1. La qualité des appariements de la région est supérieure ou égale à celles de tous ses fils. Ce qui se traduit par :

$$Q(R) \geq Q(R_f) \quad \forall R_f \in Fils(R) \quad (6.13)$$

Condition 2. Les plans 3D correspondant aux fils de la région R sont quasi-coplanaires.

$$|e(\perp \pi_f^i, \perp \pi_f^j)| < \epsilon \quad i \neq j \quad \text{et} \quad \pi_f \in \Pi\{R_f, \forall R_f \in Fils(R)\} \quad (6.14)$$

avec

- $\Pi\{R_f\}$ la famille des hypothèses de plans correspondant à R_f .
- $\perp \pi_f^i$ la normale au plan π^i de la région R_f .
- $e(\perp \pi_f^i, \perp \pi_f^j)$ est l'écart angulaire entre les normales des plans.

Le choix d'un seuil sur la qualité des appariements d'une région étant critique, la coupe se fait par un critère relatif où l'on compare la qualité de la région par rapport à celles de ses fils. Si la région mère a une meilleure qualité que tous ses fils, il faut vérifier que les fils renvoient des plans quasi-coplanaires ce qui veut dire qu'ils constituent une sur-segmentation de la facette. Dans ce cas, la région mère sera retenue dans la segmentation optimale.

Il faut remarquer que la qualité de la coupe dépend fortement de l'estimation des plans 3D. On peut aboutir à des coupes sursegmentées si les plans 3D sont mal estimés. En effet, pour un fils donné, une erreur d'estimation de la pente d'un plan entraîne un grand écart angulaire entre ses plans, d'où la descente à un sous-niveau de la hiérarchie et la sur-segmentation. Cet inconvénient peut être résolu par une étape de fusion des plans de la scène 3D finale.

6.5.5.2 Validation des appariements

Les deux segmentations optimales dans les hiérarchies fournissent des correspondances 1 à N . La validation des appariements sera réalisée en fonction de l'application visée. Pour avoir des hypothèses de facettes uniques par région planimétrique, un ensemble d'appariements consistant est retenu par hiérarchie suivant le schéma de "the winner takes all" en privilégiant la qualité de l'appariement, ensuite le score de recouvrement et enfin le score de similarité.

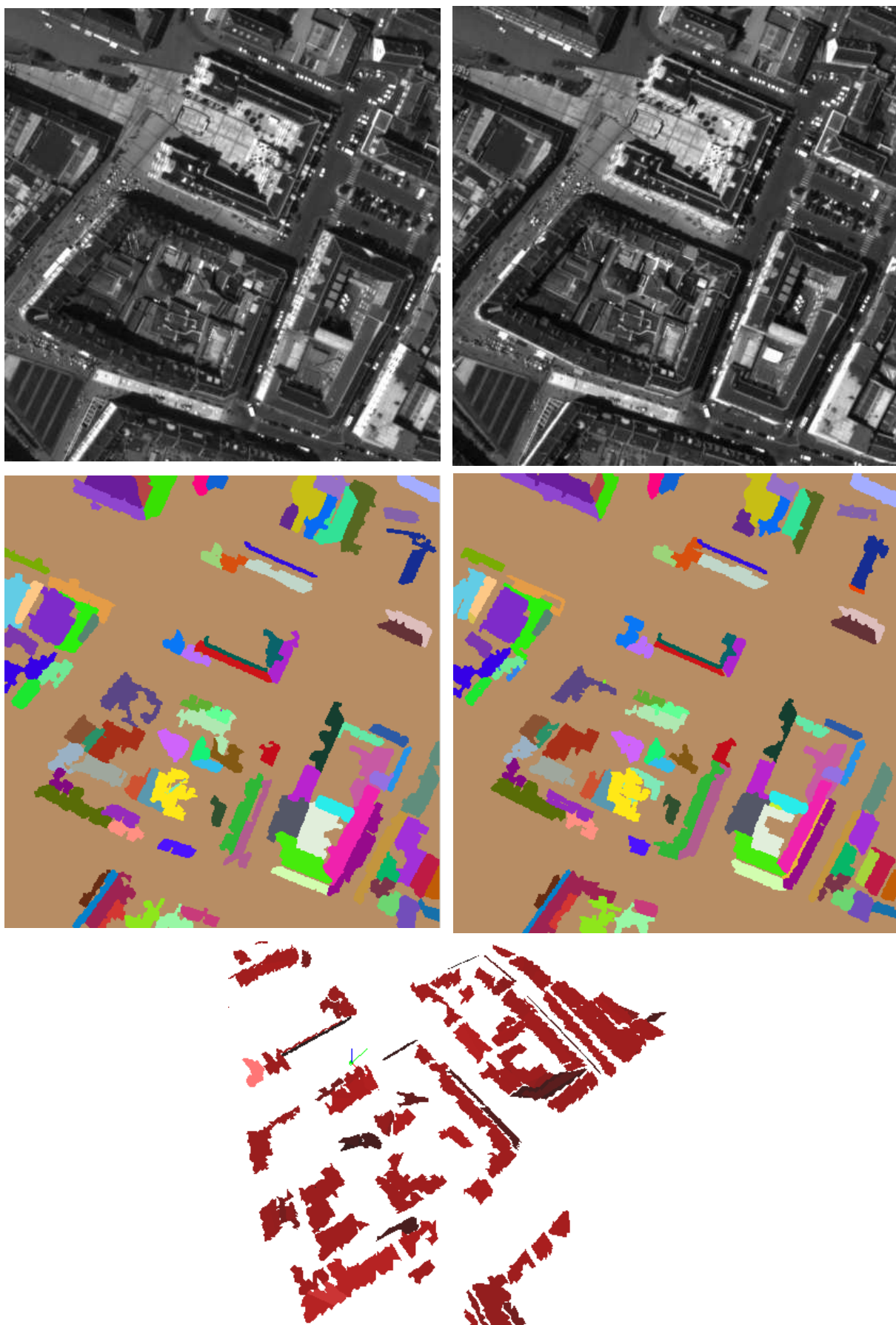
Pour avoir un ensemble d'hypothèses concurrentes de facettes 3D par région, les ensembles d'associations des deux hiérarchies sont fusionnés et élagués par la suite par le schéma de "**The winner takes all**". En fusionnant les deux segmentations, l'équation 6.12 n'est plus vérifiée, c'est-à-dire qu'une région et ses noeuds fils peuvent appartenir à l'ensemble des associations finales ce qui permet d'avoir des hypothèses de facettes concurrentes pour une même région planimétrique.

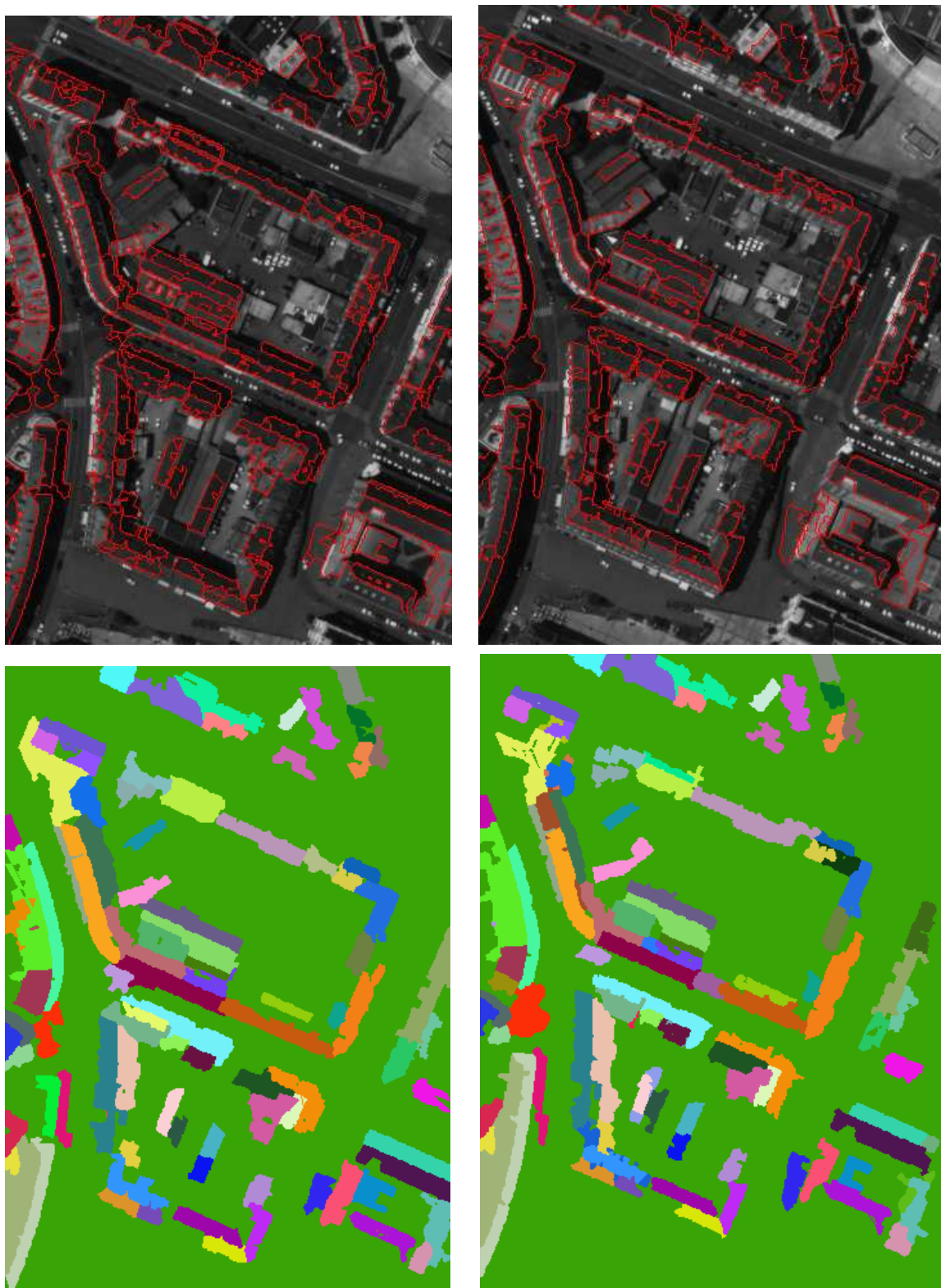
6.6 Résultats

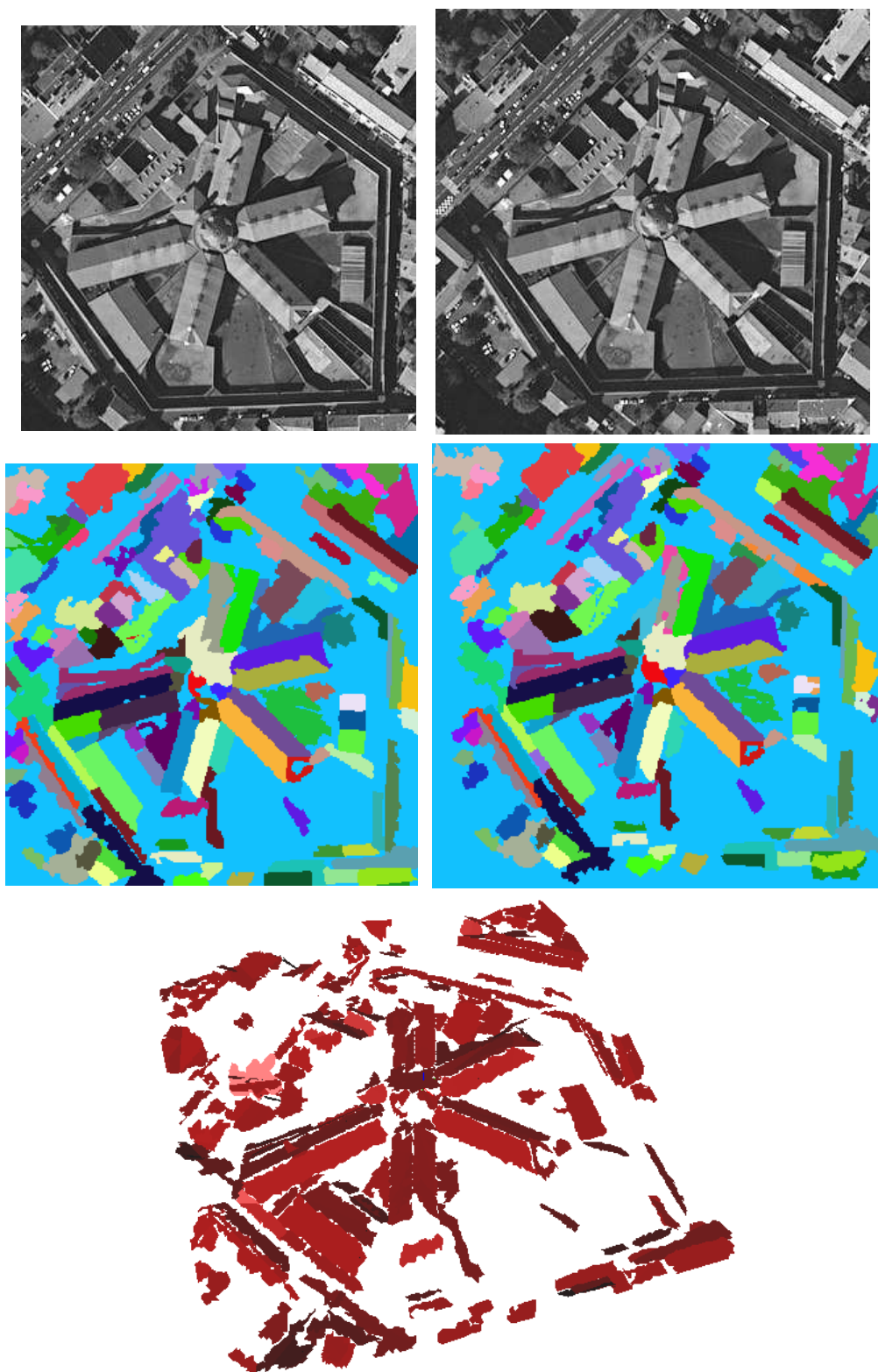
Les résultats sont présentés sur des images simulées par le CNES/QTIS² sur le site d'Amiens à 50 cm (cf. figures 6.13, 6.14), 70 cm (cf. figures 6.16, 6.17) avec un rapport $B/H = 0.2$ et sur le site de Toulouse à 60 cm (cf. figure 6.15), 80 cm (cf. figure 6.18) et un plus faible rapport $B/H = 0.08$.

Afin de se focaliser sur les bâtiments, une classification préliminaire sol/sur-sol a été effectuée sur les images pour éliminer les régions du sol.

²Centre National d'Etudes Spatiales, Toulouse
QTIS : Qualité et Traitement des Images Spatiales.

FIG. 6.13 – Amiens - 50 cm - $B/H = 0.2$

FIG. 6.14 – Amiens - 50 cm - $B/H = 0.2$

FIG. 6.15 – Toulouse - 60 cm - $B/H = 0.08$

6.6.1 Impact de la résolution

Afin de qualifier l'influence de la résolution sur l'extraction de facettes 3D, les algorithmes ont été testés sur le centre-ville d'Amiens à 70 cm avec le même rapport $B/H = 0.2$ et sur la ville de Toulouse avec une résolution de 80 cm et un $B/H = 0.08$.

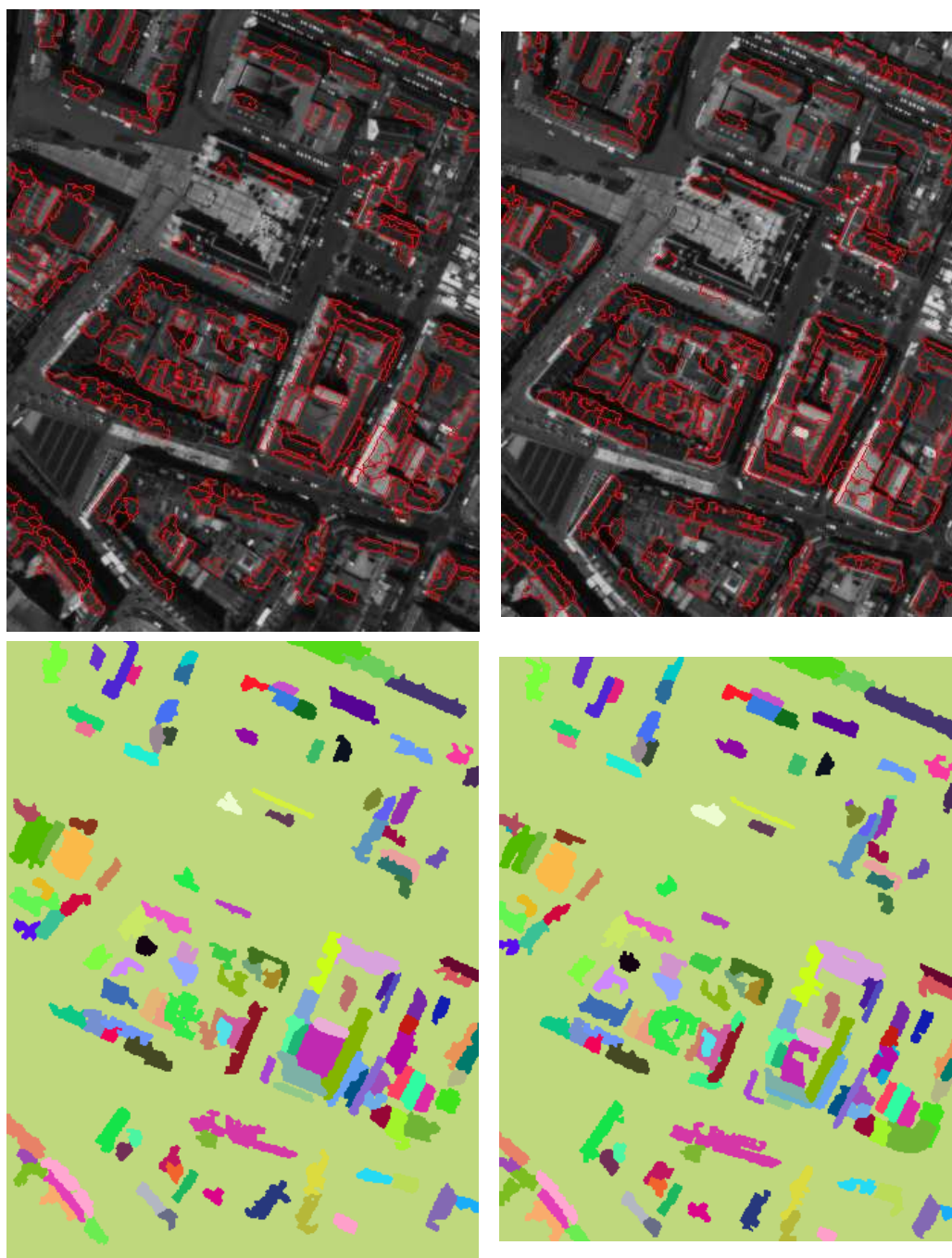
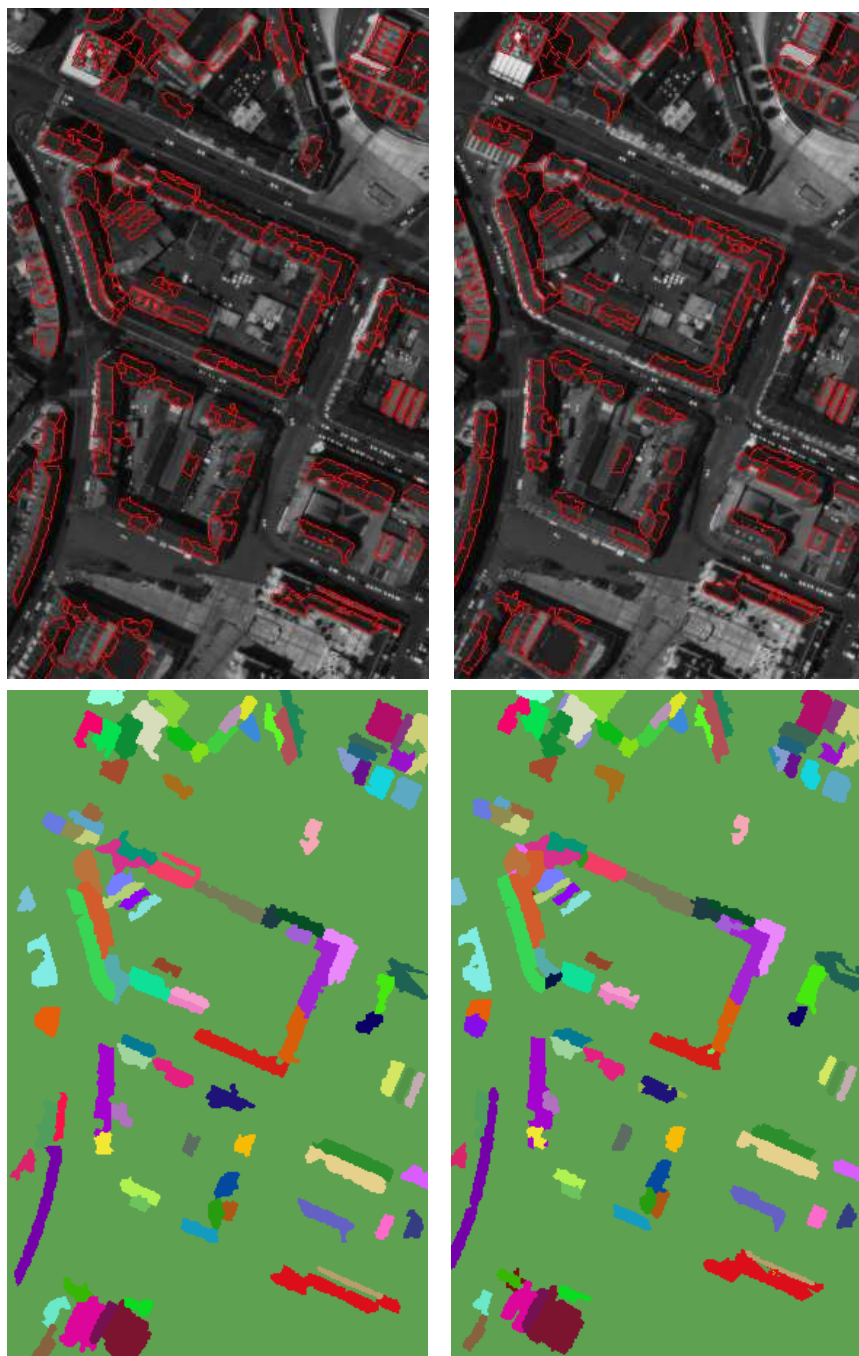
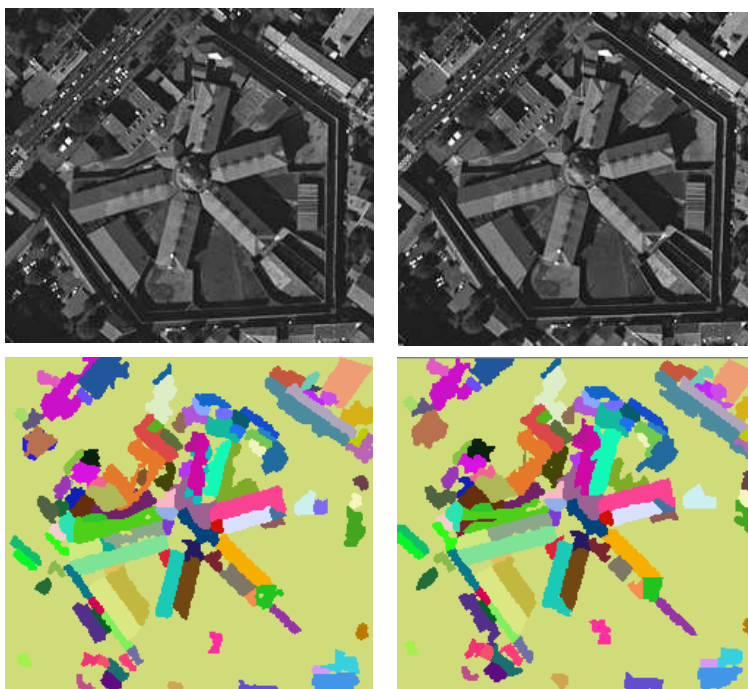


FIG. 6.16 – Amiens - 70 cm - $B/H = 0.2$

FIG. 6.17 – Amiens - 70 cm - $B/H = 0.2$

FIG. 6.18 – Toulouse - 80 cm - $B/H = 0.08$

6.7 Évaluations

Les évaluations portent sur la qualité de l'appariement entre images, sur la précision et l'exhaustivité de la reconstruction 3D. Les évaluations sont réalisées en 2D et en 3D. L'évolution des résultats en fonction de la résolution des images est également présentée.

6.7.1 Qualité de l'appariement

L'évaluation de la qualité d'appariement porte tout d'abord sur la pertinence des contraintes de similarité et de recouvrement. Dans un deuxième temps, on s'est intéressé à l'étape de recherche de segmentation optimale.

6.7.1.1 Les contraintes d'appariement

TAB. 6.1 – Contraintes d'appariement des régions

Contraintes	Similarité	Similarité + Recouvrement
% régions retenues	82%	23%
Nbre moyen homologues	3.99	2.09

Le seuil de similarité S^s n'étant pas strict, le nombre moyen d'homologues est de l'ordre de 4. Par contre la contrainte de recouvrement permet d'élaguer considérablement les associations. À l'issue des deux contraintes, uniquement 23% des régions de l'hérarchie sont appariées. Le nombre moyen d'homologues pour chaque région est de l'ordre de 2.

6.7.1.2 La recherche de segmentation optimale dans une hiérarchie

Les régions retenues dans la segmentation optimale de la hiérarchie ont été labellisées en fonction de leur nombre d'homologues. Cette évaluation a été réalisée sur la zone de test du centre ville d'Amiens à 50cm.

TAB. 6.2 – Évaluation de la coupe optimale

Régions retenues/ Régions appariées 7.5%	
Régions fiables 25%	régions à plusieurs homologues 75%

Le nombre de nœuds initiaux est de 21618. D'après le tableau 6.2, uniquement 7.5% des régions appariées sont retenues pour la segmentation optimale, dont 25% de régions fiables c'est-à-dire à un seul homologue. Pour les nœuds à plusieurs homologues, dans 97% des cas, les homologues se trouvent sur la même branche de la hiérarchie c'est-à-dire qu'ils ont une relation de dépendance : mère-fils, ce qui montre la robustesse de l'appariement. Ceci est notamment obtenu grâce à la propagation ascendante de la contrainte épipolaire qui évite d'inclure des régions parasites à un niveau plus haut dans la hiérarchie.

6.7.2 Exhaustivité de la reconstruction

L'exhaustivité de la reconstruction a été évaluée à deux niveaux. Une évaluation 2D permet de voir l'influence des seuils de taille min et max des régions : S_{min}^t, S_{max}^t sur les taux de détection et de sur-détection des facettes. Une évaluation 3D permet de mettre en évidence les hypothèses de facettes concurrentes obtenues ainsi que l'impact de la résolution de l'image sur l'exhaustivité et la précision de la reconstruction 3D. Ces évaluations ont été réalisées avec la plateforme EvalMne (cf. section 8).

6.7.2.1 Évaluations 2D

La plateforme EvalMne permet de comparer des images de facettes labellisées (cf. chapitre 8.3.2) et fournit ainsi un premier mode d'évaluation des facettes 3D extraites (cf. figure 6.19). On dispose en entrée d'un ensemble de facettes de référence projetées, à partir du modèle vecteur de référence, en utilisant l'algorithme du Z-Buffer et d'un ensemble de facettes 2D extraites. Les facettes sont appariées par un critère de recouvrement planimétrique. L'ensemble des facettes appariées est évalué et permet de calculer les taux de détection T_d et de sur-détection T_{sd} des facettes par rapport à la référence.

Sur la zone de test à 50 cm (cf. figure 6.19), le taux de détection des facettes est de 50%. Par contre, on remarque que le taux de sur-détection est très faible ce qui confirme la fiabilité des primitives extraites. On atteint bien l'objectif initial qui était de privilégier la fiabilité des facettes extraites par rapport à l'exhaustivité de la reconstruction.

Ce processus d'évaluation sera appliqué sur plusieurs zones test et permettra de mettre en évidence l'impact des différents seuils utilisés lors des contraintes.

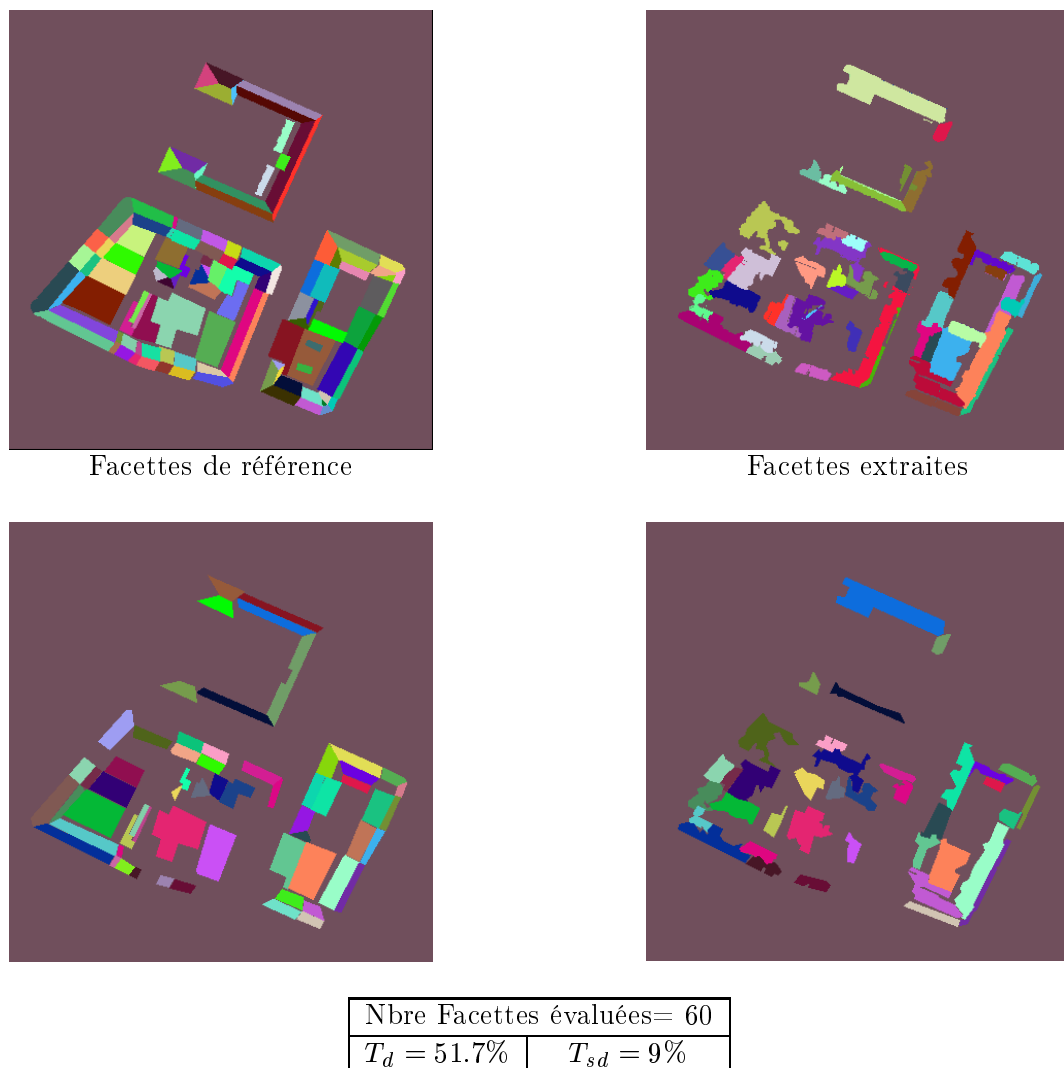


FIG. 6.19 – Procédé d'Évaluation des facettes 2D avec EvalMne, Amiens 50 cm

6.7.2.2 Influence du seuil de taille de région

L'extraction des facettes a été réalisée à 50 cm.

TAB. 6.3 – Influence de la taille maximale de région

Taille Min	Taille Max	Nb Facettes évaluées	T_d	T_{sd}
180	2000	60	51.7%	9%
180	3000	62	53%	8%
180	4000	65	56%	5.8%

Les tableaux 6.3 et 6.4 illustrent l'influence des seuils de taille minimale et maximale des régions traitées sur la qualité de la détection. Quand la taille maximale augmente, le taux de sur-détection diminue. En effet, la segmentation optimale est réalisée à un niveau plus haut dans la hiérarchie, il y a donc moins de régions sursegmentées. Quand la taille minimale augmente, le

TAB. 6.4 – Influence de la taille minimale de région

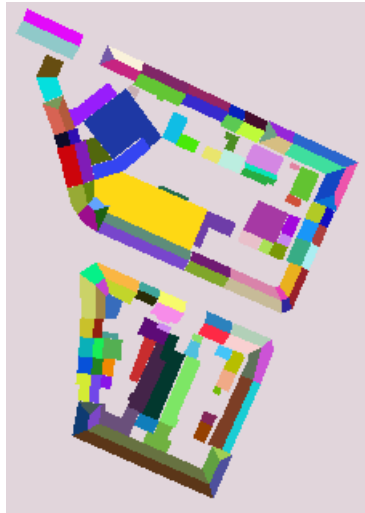
Taille Min	Taille Max	Nb Facettes évaluées	T_d	T_{sd}
120	2000	67	57.7%	10.6%
150	2000	63	54.3%	8.7%
180	2000	60	51.7%	9%

taux de sur-détection diminue également puisqu'il y a moins de sur-segmentations des régions. Dans les deux cas, on peut obtenir des sur-segmentations des facettes.

Pour notre application finale, les facettes extraites doivent être fiables. De ce fait, il faut éviter les sous segmentations de facettes qui impliquent des mauvaises estimations de plans 3D et qui risquent de fausser le résultat final. Les sur-segmentations de régions ne sont pas aussi gênantes puisqu'elle renvoient les mêmes équations de plans 3D.

Pour des images de résolution 50 cm, $S_{min}^t = 180$ correspond à une taille de région minimale de 45 m². $S_{max}^t = 2000$ correspond à une région maximale de 500 m². Ces valeurs constituent un bon compromis pour extraire les principaux pans de toits.

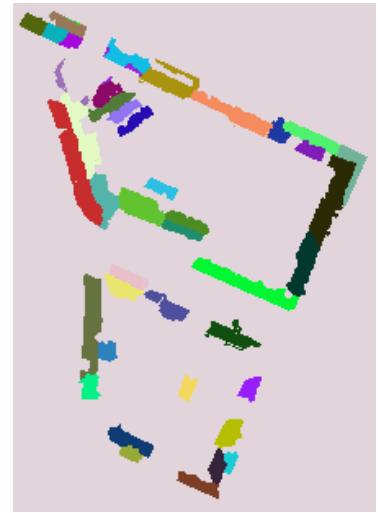
La figure 6.20 montre l'influence des seuils de taille sur une deuxième zone de test à 70 cm. Les taux de sur-détection des facettes en 2D sont faibles. Les taux varient sensiblement avec le seuil de taille minimale.



Facettes 2D de référence



Facettes extraites (70-1000)



Facettes extraites (100-1000)

Nbre Facettes évaluées = 65	
$T_d = 49\%$	$T_{sd} = 5\%$

Nbre Facettes évaluées = 61	
$T_d = 46\%$	$T_{sd} = 3\%$

FIG. 6.20 – Impact du seuil de taille à 70 cm

6.7.2.3 Évaluation en fonction de la résolution

La figure 6.21 compare deux ensembles de facettes extraites sur la même zone de test, à deux résolutions différentes : 50 et 70 cm. Les seuils de taille S_{min}^t et S_{max}^t sont adaptés à la résolution de l'image. Le taux de détection des facettes est supérieur à 50 cm, ce qui était prévisible. Par contre le taux de sur-détection des facettes reste faible aux deux résolutions.



Facettes extraites 50 cm (180-2000)

Nbre Facettes évaluées = 67	
$T_d = 51.1\%$	$T_{sd} = 8\%$



Facettes extraites 70 cm (100-1000)

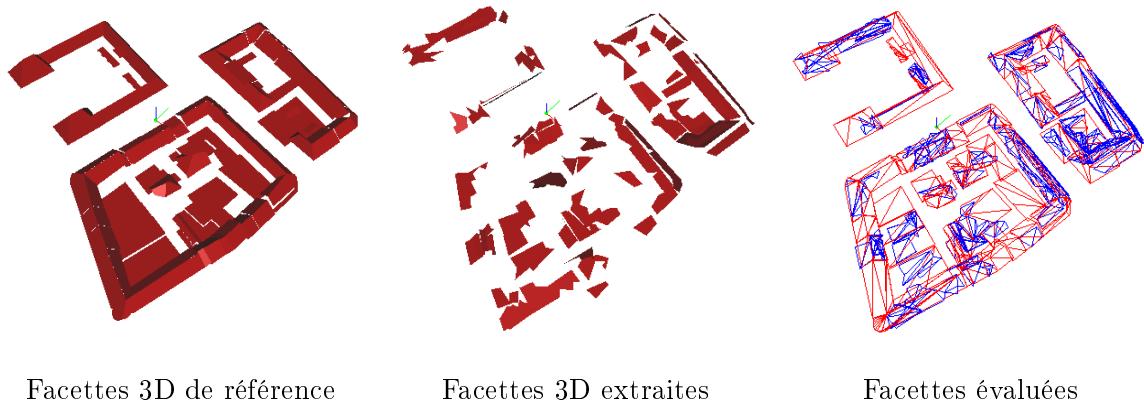
Nbre Facettes évaluées = 61	
$T_d = 46\%$	$T_{sd} = 3\%$

FIG. 6.21 – Évaluation 2D en fonction de la résolution

6.7.2.4 Évaluations 3D

Le premier niveau d'évaluation a été réalisé en 2D ou encore en raster, sur les facettes 3D projetées dans les images. Les hypothèses concurrentes de facettes n'ont pas été prises en compte à cause de l'algorithme de Z-buffer qui renvoie la facette supérieure. Dans cette partie, les évaluations sont réalisées en 3D. Les comparaisons auront lieu entre les facettes 3D extraites et les modèles vectoriels de référence. L'avantage de l'évaluation en 3D est de pouvoir prendre en compte l'ensemble des facettes concurrentes pour une même région planimétrique et de pouvoir évaluer la précision de la reconstruction 3D.

On a utilisé le même processus d'évaluation que pour les facettes basées sur les segments 3D (cf. figure 6.22). La figure de droite illustre la superposition des facettes appariées entre le modèle de référence (en rouge) et les facettes extraites (en bleu). Le tableau indique les taux de détection et sur-détection des facettes ainsi que la moyenne de l'erreur angulaire et son écart-type par rapport aux plans des facettes de référence.



Nbre Facettes évaluées = 244	
$T_d = 80\%$	$T_{sd} = 11\%$
$\bar{\epsilon} = 15.4^\circ$	$\sigma_\epsilon = 10^\circ$

FIG. 6.22 – Évaluation 3D des facettes

Si l'on compare les résultats des évaluations 2D (cf. figure 6.19) et 3D (cf. figure 6.22) sur la même zone de test, on remarque tout de suite la différence du nombre de facettes évaluées, 75% des facettes ne sont pas évaluées en 2D. Ceci est du principalement à l'étape de projection dans les images qui ne permet pas de traiter les facettes concurrentes pour une même région planimétrique, ainsi qu'aux parties cachées.

En 3D, on obtient un bon taux de détection des facettes 3D sur ces bâtiments qui est de 80%. Le taux de sur-détection des facettes est de seulement 11%. Les résultats confirment la fiabilité des facettes extraites par approche région, contrairement à l'approche basée sur les segments où le taux de sur-détection des facettes est trop élevé pour pouvoir les utiliser dans le processus d'optimisation globale. On peut remarquer toutefois que le taux de sur-détection est supérieur en 3D à cause des hypothèses de facettes concurrentes pour une même région planimétrique.

6.7.2.5 Impact de la résolution

La figure 6.23 illustre l'impact de la résolution sur les taux de détection et sur-détection ainsi que sur la précision de la reconstruction. On a utilisé une deuxième zone de test sur Amiens à 50 et 70 cm. Les seuils de taille minimale et maximale sont adaptés aux résolutions :

à 50 cm, $(S_{min}^t, S_{max}^t) = (180, 2000)$.

à 70 cm, $(S_{min}^t, S_{max}^t) = (100, 1000)$.

Le taux de détection des facettes 3D à 50 cm est nettement supérieur que celui obtenu à 70 cm, ce qui a été également vérifié en 2D. Par contre les taux de sur-détection sont comparables et restent faibles.

6.7.3 Temps de calcul

Les temps de calcul sont donnés sur la zone de test du centre ville d'Amiens à 50 cm. La taille de l'image est de 411*427 pixels soit 4,38ha. On utilise comme seuils de taille $(S_{min}^t, S_{max}^t) = (180, 2000)$.

Un masque du sol est extrait en semi-automatique par croissance de régions à partir du MNE

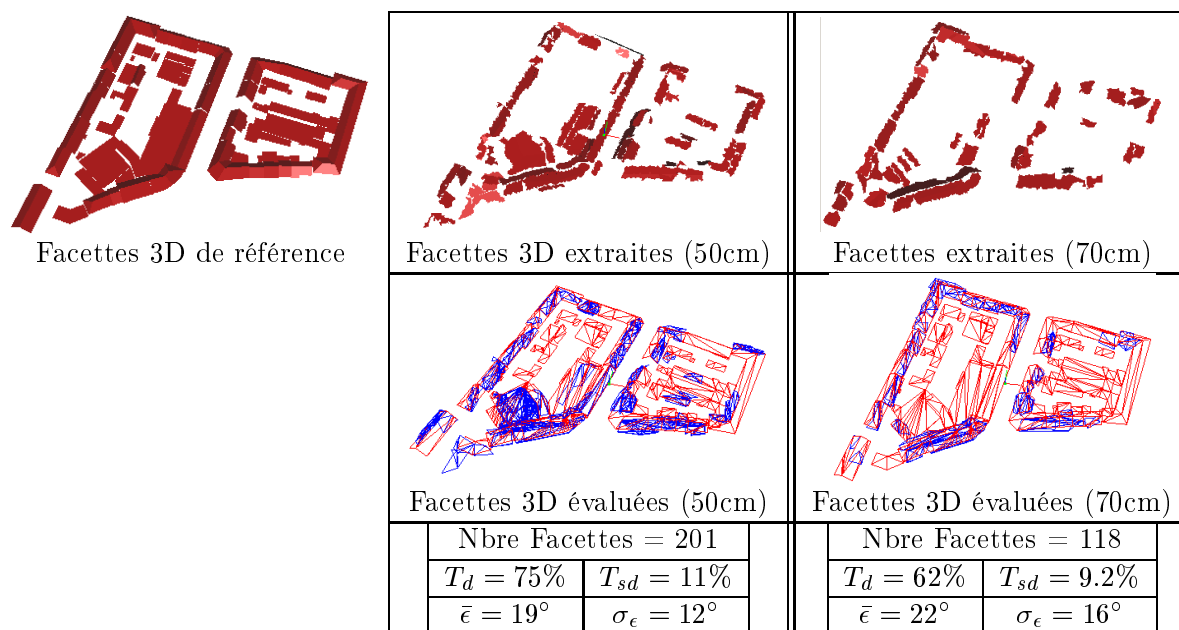


FIG. 6.23 – Évaluations 3D : Impact de la résolution

initial et permet le marquage des nœuds de base appartenant au sol.

- Pour une hiérarchie, le nombre de nœuds initiaux = 21618 (sol : 10157 nœuds, bâtiments : 11461 nœuds).
- Application de la contrainte épipolaire sur les nœuds de base : 6757 nœuds : 71 s
- Propagation ascendante de la contrainte épipolaire : 21595 nœuds : 4.36 s.
- Appariement et validation par approche descendante : 1609 nœuds traités : 46.8 s.
- Ensemble des régions retenues pour la segmentation optimale : 121 nœuds.

On remarque que l'étape d'application de la contrainte épipolaire sur les nœuds de base nécessite beaucoup de temps de calcul par rapport aux autres contraintes. La propagation ascendante de la contrainte permet un gain de temps considérable.

En pratique, les contraintes de similarité et de recouvrement sont appliquées par une stratégie descendante et permettent de piloter la recherche de la segmentation optimale de la hiérarchie. Cette étape nous évite d'appliquer les contraintes sur tous les nœuds de la hiérarchie. Uniquement 14% des nœuds "bâtiments" sont traités lors de la recherche de la segmentation optimale. Les régions retenues pour la segmentation optimale, constituent 7.5% de l'ensemble des nœuds traités dans l'approche descendante.

6.8 Conclusions

On a présenté dans ce chapitre une méthode de détection et de reconstruction de facettes 3D adaptée au contexte satellitaire. La méthode est basée sur la mise en correspondance de régions entre les deux images afin d'émettre des hypothèses de facettes 3D. Les algorithmes mis en oeuvre mettent à profit le faible rapport B/H du couple stéréoscopique. Ce paramètre a influencé le choix des méthodologies. En effet, la similarité des deux segmentations nous a conduit à choisir l'appariement de régions dans l'espace image. D'autre part, à des résolutions

submétriques, les régions constituent des primitives plus fiables que les primitives linéiques et fournissent une meilleure estimation des plans 3D.

6.8.1 Apports

La méthode de reconstruction de facettes 3D proposée se base sur trois grandes parties classiques. Tout d'abord, la segmentation des images en régions. On utilise une description multi-échelles qui fournit plusieurs niveaux d'interprétation de la scène. Cette description est adaptée au contexte satellitaire puisqu'elle permet plusieurs niveaux de description de la scène et de généralisation des bâtiments.

La deuxième étape est la mise en correspondance des régions. La nouveauté, dans notre approche, est l'appariement global des deux segmentations multi-échelles. Comme stratégie d'appariement, on utilise une propagation ascendante des contraintes dans la hiérarchie, notamment la contrainte épipolaire et la contrainte d'adjacence, qui rend l'algorithme plus robuste au bruit, aux sur-segmentations des régions et réduit considérablement le temps de calcul. L'appariement des régions est basé sur des critères géométriques et fournit des associations à différents niveaux de détail. Les deux images sont traitées de façon symétrique. La validation des appariements est réalisée par une stratégie descendante (**Coarse To Fine**) puisque les appariements sont plus fiables sur des plus grandes régions. La validation est pilotée par la mise en correspondance des régions et combine par conséquent des contraintes photométriques (scores de corrélation entre régions) et des contraintes de planéité des régions (surfaces 3D reconstruites à partir des régions appariées). L'appariement global des hiérarchies renvoie deux segmentations non denses des images qui sont appariées entre elles. Les facettes 3D correspondantes sont reconstruites directement.

En fonction de l'application visée, on a la possibilité d'avoir des hypothèses de facettes 3D uniques ou concurrentes par région planimétrique, soit en utilisant une contrainte d'unicité entre les deux ensembles de régions, soit en fusionnant toutes les associations et en ne gardant que les meilleures par le schéma de "The winner takes all".

L'approche fournit de bons taux de détection des facettes 3D et surtout un faible taux de sur-détection. Ces facettes 3D constituent des primitives 3D fiables que l'on pourra utiliser dans le processus de modélisation globale de la scène (cf. section 7.3.4).

6.8.2 Inconvénients

L'inconvénient de cette approche est incontestablement le temps de calcul que nécessite les segmentations multi-échelles, le calcul du cube de corrélation ou encore l'application de la contrainte épipolaire sur les nœuds de base. Cependant le temps de calcul n'était pas un paramètre critique dans les objectifs fixés.

Le deuxième inconvénient apparaît dans la recherche de la segmentation optimale. La méthodologie adoptée tend à renvoyer des sur-segmentations de régions. En effet, l'approche est basée sur la planéité des régions, or sur des petites régions, l'estimation des plans n'est plus fiable. Pour pallier à ce problème, on introduit un seuil de taille minimale qui règle le niveau d'interprétation voulu.

6.9 Perspectives

L'appariement global des deux hiérarchies est réalisée dans notre cas, en deux temps. On extrait tout d'abord, un ensemble de régions par hiérarchie, qui représente une segmentation non

dense de la scène et qui renvoie les principales facettes 2D. Dans un deuxième temps, une fusion entre les deux ensembles de régions permet d'obtenir les associations finales. On peut envisager un processus qui réalise conjointement les segmentations dans les deux hiérarchies et leur mise en correspondance. Ce problème pourrait être réalisé par un couplage hiérarchique de graphe. La recherche de la clique maximale dans le graphe renvoie deux partitions des deux images appariées entre-elles. Cette approche pourrait être discutée, notamment pour les parties cachées qui ne peuvent pas être appariées. Le fait d'imposer des partitions denses des images, soient appariées entre elles, risque de fausser les résultats et les appariements obtenus ne seraient pas forcément fiables.

Pour améliorer la précision de reconstruction des facettes 3D, on envisage également une étape de recalage fin des facettes a posteriori. Le cube de corrélation sera discrétisé plus finement en Z . Le recalage des plans des facettes sera basé sur la maximisation des scores de corrélation sur les voxels de la facette 3D. Pour la mise en œuvre, on peut utiliser un algorithme d'optimisation contrainte tel que Levenberg-Marquardt où les points 3D seraient pondérés par leurs scores de corrélation dans le cube.

6.10 Bilan de la reconstruction de primitives

On a présenté dans les chapitres (4,5,6) une approche de reconstruction de segments 3D et deux méthodes de reconstruction de facettes 3D, la première est basée sur les segments 3D extraits et la seconde utilise une approche régions. Les segments 3D présentent l'avantage d'avoir une bonne localisation planimétrique qu'on mettra à profit, par contre ils souffrent d'une forte imprécision altimétrique. Pour les facettes 3D, on retiendra essentiellement la deuxième approche qui fournit des facettes 3D fiables avec un faible taux de sur-détection. Cependant, comme prévu, les détecteurs de primitives souffrent de problèmes de sous détection et sur-détection. La description de la scène n'est pas exhaustive avec les primitives extraites. On peut également avoir plusieurs hypothèses concurrentes pour une même région planimétrique. À partir de ces données, il est nécessaire de réaliser une optimisation globale de la surface 3D en exploitant au mieux les primitives 3D extraites. Cette étape sera détaillée dans la suite du manuscrit.

Quatrième partie

Modélisation de surfaces 3D

Chapitre 7

Modélisation de surfaces 3D

7.1 Introduction

Les processus d'extraction de primitives 3D détaillés dans la partie III fournissent un ensemble de primitives 3D fiables et complémentaires. Cependant, cet ensemble souffre de sous-détections de primitives c'est-à-dire que la simple description de la scène avec ces primitives ne sera pas dense. Il est, par conséquent, indispensable de procéder à une modélisation globale de la scène qui prend en compte les primitives extraites et qui devra tirer profit de leur complémentarité. La modélisation globale permettra également de résoudre le problème de sur-détection de primitives puisqu'une seule hypothèse sera validée parmi toutes les hypothèses concurrentes de facettes 3D. Les différentes hypothèses peuvent provenir soit d'un même processus (cf. chapitre 6) soit de plusieurs processus. Notre objectif est d'obtenir un modèle numérique d'élévation hybride (raster/vecteur) qui correspond à une description 3D dense de la scène.

La modélisation de la surface 3D revient à un problème d'appariement entre les images. Dans la section 2.3, un état de l'art détaillé a été présenté sur les approches de mise en correspondance de deux images. Trois approches se distinguent : les méthodes surfaciques qu'elles soient locales ou globales, les méthodes à base de primitives et enfin les méthodes énergétiques dont les approches basées sur la minimisation d'énergie. Dans notre stratégie, on a choisi la minimisation d'énergie car elle permet une intégration et une sélection naturelle de différentes primitives 3D (points, segments, facettes) dans un modèle de surface. Dans ce chapitre, on traitera le problème de minimisation d'énergie pour la reconstruction d'une surface 3D à partir d'un couple d'images. On introduit ainsi le processus d'optimisation 3D contrainte de la scène qui permet d'aborder le problème de modélisation, fixé par cette thèse, pour la génération d'un MNE hybride raster/vecteur.

7.2 Minimisation d'énergie : État de l'art

De nombreuses publications en vision proposent des algorithmes de minimisation d'énergie pour différentes applications telles que la segmentation d'images [IG98b, Vek00, BJ01, BT99], la stéréoscopie [RC98, IG98a, BVZ98, KZ01], la reconstruction d'objets [SVZ00], la reconnaissance d'objets [BH99] ou encore la réalité virtuelle [TBRN00]. Dans cette section, on s'intéressera au problème de minimisation d'énergie pour la reconstruction d'une surface 3D à partir d'un couple d'images. La modélisation est cependant suffisamment générique pour prendre en compte un nombre quelconque d'images.

La modélisation de surfaces 3D revient à un problème d'appariement entre images, autrement dit à une estimation de la disparité. Ce problème peut être formulé comme un problème de minimisation d'énergie où la fonction d'énergie est définie comme suit :

$$E(d) = E_{donnees}(d) + \lambda E_{lissage}(d) \quad (7.1)$$

Le but étant de trouver la fonction de disparité d qui minimise cette énergie. Entre deux pixels homologues p_1 et p_2 , la disparité $d = disp(p_1, p_2)$ est la variation de position entre les deux pixels. Elle peut être également interprétée comme une profondeur par rapport au système de prise de vue (cf. section 2.2.2.3). λ est un terme de régularisation qui donne plus ou moins de poids à l'énergie de lissage $E_{lissage}$.

$E_{donnees}(d)$ correspond au terme d'attache aux données et mesure à quel point la disparité respecte les données images initiales.

$$E_{donnees}(d) = \sum_{x,y} C(x, y, d(x, y)) \quad (7.2)$$

x, y étant les coordonnées des pixels de l'image et C la fonction de coût d'appariement.

$E_{lissage}(d)$ correspond au terme d'attache au modèle global de la scène. C'est une énergie de lissage qui tend à pénaliser les différences de disparités entre pixels voisins.

$$E_{lissage}(d) = \sum_{x,y} \rho(d(x, y) - d(x + 1, y)) + \rho(d(x, y) - d(x, y + 1)) \quad (7.3)$$

où ρ est une fonction croissante de d . Certains algorithmes de régularisation de surfaces [PTC85] utilisent des fonctions ρ quadratiques, ce qui tend à lisser la disparité sur toute l'image et renvoie par conséquent de mauvais résultats sur les bords des objets. D'autres fonctions d'énergie préservent les discontinuités grâce à des fonctions ρ robustes [Ter86].

Le terme E_{smooth} peut aussi dépendre de la différence d'intensité, e.g

$$\rho_d(d(x, y) - d(x + 1, y)) \cdot \rho_I(\| I(x, y) - I(x + 1, y) \|) \quad (7.4)$$

Cette idée développée dans [BVZ99, Fua93] favorise les discontinuités de disparité en présence de contours (intensité).

Une fois la fonction d'énergie définie, il existe plusieurs méthodes pour la minimiser, le minimum pouvant être local. Dans [SSZ02], les auteurs passent en revue les méthodes d'optimisation globale. On distingue deux groupes de minimisation d'énergie : le premier groupe est basé sur la programmation dynamique où l'on cherche le minimum indépendamment pour chaque ligne [Bel96, GLY95, BI99]. Le deuxième repose sur des formulations par graphes où l'on cherche à minimiser directement en 2D la surface de disparité [RC98, IG98a, BVZ99, Vek99, KZ01]. Ces algorithmes sont basés sur les coupes de capacité minimale ou de façon duale la recherche de flot maximal dans un graphe. Ces techniques ne sont valides que pour des formulations particulières de l'énergie à minimiser [KZ02]. Elles sont toutefois valides pour une grande famille d'énergies utilisées pour la reconstruction de surfaces.

7.2.1 Formulation du problème par programmation dynamique

Le problème d'optimisation consiste à trouver le chemin de coût minimal dans une matrice de coûts d'appariement de tous les points de deux lignes correspondantes dans les deux images. Les occultations sont modélisées par l'appariement d'un ensemble de pixels d'une image à un pixel unique dans la seconde image et la solution est alors pénalisée par un coût d'occultation. La

programmation dynamique renvoie le minimum global en un temps de calcul linéaire en fonction du nombre de pixels mais, pour chaque ligne séparément. Les problèmes de la programmation dynamique, sont principalement la sélection du coût d'occultation, la difficulté à rajouter des consistances inter-lignes, ce dernier problème a été traité dans [OK85, BI99] et la nécessité de la contrainte d'ordre.

Geiger et al. [GLY95] proposent une méthode basée sur une approche bayésienne utilisant les contraintes d'ordre, d'unicité et d'occultation. Les auteurs font l'observation que sur une ligne épipolaire, une discontinuité de disparité dans une image correspond toujours à une partie cachée dans la deuxième image et vice versa. La fonction de disparité optimale est obtenue par estimation du maximum a posteriori, en utilisant la programmation dynamique.

Belhumeur [Bel96] a développé un modèle basé sur une approche bayésienne où les occultations sont représentées explicitement. Il tient compte simultanément de la profondeur, des parties cachées, de l'orientation et des irrégularités de surface.

7.2.2 Formulation du problème par flot maximal

Le problème de flot de graphe est un problème classique en algorithmique. Il peut être formulé comme un problème d'écoulement d'eau dans un réseau de tuyaux. On explique dans un premier temps le problème de manière intuitive [PSQ04] avant de présenter la formulation théorique.

7.2.2.1 Écoulement d'eau dans un réseau

Étant donné une source d'eau de débit infini, un puits de contenance infinie et un réseau de tuyaux reliant la source au puits, on cherche le flot maximal que l'on peut faire passer dans le réseau. Comme le débit de la source et la contenance du puits sont infinis, le flot maximal est uniquement contraint par le réseau. Le réseau constitue, par conséquent, un goulot d'étranglement. En effet, pour aller de la source au puits, le flot doit emprunter l'un de ces tuyaux. Donc, dans le meilleur des cas, si les tuyaux sont pleins d'eau, le flot est maximal et est égal à la somme de leurs capacités. Le goulot d'étranglement est alors constitué d'un ensemble de tuyaux de capacité minimale dont la somme renvoie la valeur du flot maximal.

7.2.2.2 Formulation théorique

On considère un graphe pondéré $\mathcal{G} = \langle \nu, \mathcal{E} \rangle$, ν étant l'ensemble des nœuds et \mathcal{E} l'ensemble des arêtes orientées $\mathcal{E} \in \nu^2$, avec deux nœuds terminaux $\{s, t\}$, la source et le puits. Pour un nœud x , on définit l'ensemble des arêtes entrantes $\mathcal{E}_e(x)$ et celui des arêtes sortantes $\mathcal{E}_s(x)$:

$$\mathcal{E}_e(x) = \{\epsilon \in \mathcal{E} / \exists y \in \nu, \epsilon = (y, x)\}$$

$$\mathcal{E}_s(x) = \{\epsilon \in \mathcal{E} / \exists y \in \nu, \epsilon = (x, y)\}$$

On définit une fonction *capacité* qui associe à une arête ϵ un réel positif $C(\epsilon) = |\epsilon|$ et une fonction *flot*, $Flot(\epsilon)$. Le flot est *valide* si

$$\forall \epsilon \in \mathcal{E} \quad Flot(\epsilon) \leq |\epsilon| \quad (7.5)$$

$$\forall x \in \nu \setminus \{s, t\} \quad \sum_{\epsilon \in \mathcal{E}_e(x)} Flot(\epsilon) = \sum_{\epsilon \in \mathcal{E}_s(x)} Flot(\epsilon) \quad (7.6)$$

s et t étant exclus de la contrainte 7.6 pour simplifier la suite des définitions. On suppose que $\epsilon_s(s) = \epsilon_e(t) = \emptyset$.

À une fonction de flot donnée, on associe une valeur appelée *flot de graphe*, définie comme suit :

$$Flot(\mathcal{G}) = \sum_{\epsilon \in \mathcal{E}_s(s)} Flot(\epsilon) \quad (7.7)$$

On définit la coupe $\mathcal{C}_{\mathcal{G}}$ de \mathcal{G} , que l'on note aussi $\mathcal{C}(\nu^s, \nu^t)$, comme une partition des nœuds en deux ensembles tel que $s \in \nu^s$ et $t \in \nu^t$ (cf. figure 7.1).

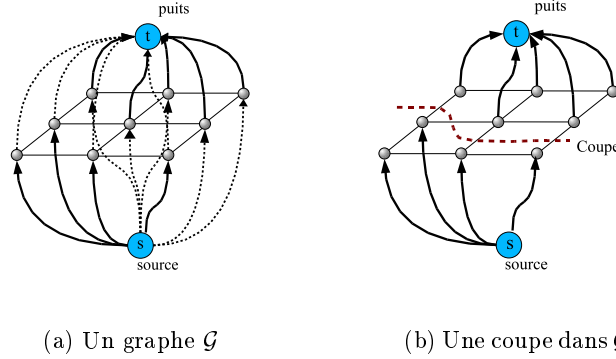


FIG. 7.1 – Exemple d'un graphe 2D orienté, pondéré. Les arêtes de capacité maximale sont illustrées en gras

La capacité d'une coupe de graphe, notée $|\mathcal{C}_{\mathcal{G}}|$ est égale à la somme des poids des arêtes coupées et qui séparent ν^s et ν^t (Eq. 7.15).

$$|\mathcal{C}_{\mathcal{G}}| = \sum_{(x,y) \in (\nu^s, \nu^t)} |\epsilon_{xy}| \quad (7.8)$$

Une arête $\epsilon(x, y)$ est **coupée** si $(x, y) \in (\nu^s, \nu^t)$.

Le problème de coupe minimale consiste à trouver la coupe de capacité minimale. Ce problème peut être résolu en calculant le flot maximal entre la source et le puits [FF62]. On a la relation suivante entre le flot maximal atteint sur l'ensemble des fonctions de flots valides et la coupe de capacité minimale.

$$\text{Max}(\text{Flot}(\mathcal{G})) = \text{Min}(|\mathcal{C}_{\mathcal{G}}|) \quad (7.9)$$

Ce qui revient à dire qu'une recherche de flot maximal correspond à une coupe de capacité minimale.

7.2.2.3 État de l'art

La première formulation de flot maximal pour la stéréovision a été proposée par Roy et Cox [RC98]. Basée sur une formulation 1-D de la contrainte d'ordre en utilisant les méthodes de programmation dynamique, ils proposent une formulation plus générale en 2D. Leur approche utilise une contrainte de cohérence locale qui suppose que la carte de disparité est lisse localement. Ils considèrent par conséquent un voisinage 6-connexe et supposent que la disparité est similaire dans toutes les directions. L'avantage de cette approche est qu'elle prend en compte les pixels voisins le long des lignes épipolaires, mais également entre les lignes épipolaires. Le graphe est construit de manière classique comme dans la plupart des problèmes de flot maximal (figure 7.2) : un maillage 3D de points (x, y, d) , où (x, y) sont les coordonnées images et d l'éventuelle disparité correspondante, auquel on rajoute la source et le puits. Le maillage est 6-connexe, la source est connectée au plan de disparité minimale et le puits est connecté au plan de disparité maximale. La carte de disparité est construite à partir de la coupe de capacité minimale : pour chaque point (x, y) , la plus grande disparité associée à l'arête correspondante est retenue.

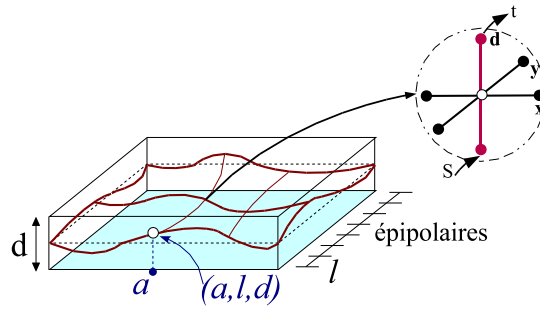


FIG. 7.2 – Construction du graphe 3D : Formulation de Roy et Cox

L'inconvénient de cet algorithme [IG98a] est qu'il ne garantit pas les contraintes d'ordre et d'unicité. De plus, les occultations et les discontinuités ne sont pas modélisées et les régions cachées sont forcément appariées. D'autre part, la formulation du flot maximal proposée par Roy et Cox ne permet de traiter que les fonctions convexes et ne peut modéliser que des pénalités linéaires pour les discontinuités.

Ishikawa et Geiger [IG98a] proposent un autre algorithme de flot maximal basé sur l'algorithme de Roy et Cox. La méthode assure les contraintes d'ordre et d'unicité. Pour modéliser explicitement les occultations et les discontinuités, le modèle inclut des contraintes géométriques qui nécessitent une correspondance mutuelle entre une discontinuité de disparité le long d'une ligne épipolaire dans une image et une partie cachée correspondante dans la deuxième image. La construction du graphe est modifiée par rapport à la construction générale : un maillage 3D de couples de points appariés, dédoublés (x, y, r) . En supposant la rectification des images en épipolaire, (x, r) sont les coordonnées du pixel dans l'image de gauche et (y, r) les coordonnées du point homologue. Les arêtes sont orientées et leurs coûts représentent les différents termes d'attache au modèle : le score d'appariement, le coût d'occultation, le coût de lissage et les contraintes géométriques. Le graphe orienté, contrairement à celui utilisé dans [RC98], permet d'assurer les contraintes d'unicité et d'ordre. Le modèle donne plus de poids à l'énergie de lissage le long des lignes épipolaires. Leur modèle suppose que les discontinuités de disparité sont liées à la présence de contours et des jonctions. Des forts gradients dans une image, réduisent les coefficients de lissage et avantagent les discontinuités. Les toits à fortes pentes sont traités séparément puisqu'ils induisent des appariements multiples, certaines arêtes sont forcées à l'infini pour respecter la contrainte d'unicité. La carte de disparité finale est obtenue par coupe minimale dans le graphe qui renvoie un ensemble d'arêtes d'appariement.

L'algorithme traite les images de façon symétrique et utilise les coupes de graphes pour gérer explicitement les discontinuités et forcer l'unicité. La solution globale est calculée en un temps polynomial. Le point faible de cette approche réside dans le terme de lissage ; la distance L_1 n'est pas robuste.

Boykov et al. [BVZ98] ont proposé une autre formulation du problème de flot maximal sur des graphes orientés. Les auteurs montrent que le problème d'appariement peut être formulé comme un champ de Markov aléatoire (Markov Random Fields MRF). Le MRF utilisé est une généralisation de la fonction d'énergie de Potts initialement décrite dans [Pot52]. Ils montrent aussi que l'estimation du Maximum a Posteriori (MAP) d'un tel MRF peut être résolue avec la recherche d'une coupe minimale "multi-way minimum cut". Boykov et al. généralisent le mo-

dèle de Potts afin d'avoir une fonction modélisant les pénalités entre pixels voisins plutôt qu'une constante. L'avantage de cette nouvelle formulation est qu'elle permet de gérer des pénalités non linéaires pour les discontinuités et donc d'avoir des cartes de disparité plus précises aux bords des objets.

Dans [BVZ99], les mêmes auteurs développent encore plus cet algorithme et insistent sur la recherche de la coupe de capacité minimale. Le graphe est construit de manière générale : les nœuds correspondent aux pixels de l'image et l'ensemble de disparités possibles. Les relations de voisinage sont respectées par des arêtes "n-links". Les coûts affectés aux arêtes individuelles correspondent aux différents termes de la fonction d'énergie. Le but étant toujours le même, de trouver la coupe de capacité minimale. Le problème étant NP-complet pour la plupart des classes de fonctions [KZ01], les auteurs proposent un algorithme pour trouver une solution approchée. La solution est recherchée de façon itérative, à chaque itération, seulement deux labels sont examinés. La première opération est le $\alpha - \beta$ *swap move* où seuls les pixels affectés des labels α et β sont pris en compte et sont réassignés de façon à diminuer l'énergie globale. La deuxième opération est le $\alpha -$ *expansion move*, où le label α est assigné autant de fois que possible tout en diminuant l'énergie globale. La nouvelle labellisation est trouvée grâce à la coupe de capacité minimale.

Dans [KZ01], les auteurs proposent un algorithme qui s'inscrit dans la suite de celui de Boykov et qui traite de façon symétrique les deux images. Pour cela, ils assignent des labels binaires à des paires de pixels provenant des deux images, au lieu d'assigner des labels individuels aux pixels d'une image de référence. La fonction d'énergie inclut de manière explicite une énergie d'occultation qui force la contrainte d'unicité (cf. eq 7.10).

$$E(f) = E_{donnees}(f) + E_{occultation}(f) + E_{lissage}(f) \quad (7.10)$$

avec $E_{donnees}(f)$ la différence des intensités entre les pixels homologues.

$E_{occultation}(f)$ impose une pénalité si le point est caché.

$E_{lissage}(f)$ énergie de lissage, tend à propager la même disparité aux pixels voisins.

Le minimum approché est trouvé grâce à l'opération $\alpha -$ *expansion move*.

Un article récent de Boykov et Komogorov [BK04] présente tous les algorithmes basés sur des coupes minimales dans les graphes. Une étude expérimentale permet de comparer les différentes méthodes en terme de complexité de temps de calcul, de qualité des résultats, et ce, pour différentes applications notamment la restauration d'images, la stéréoscopie ou encore la segmentation d'images.

7.2.3 Méthodes coopératives

Des méthodes coopératives sont proposées pour résoudre le problème de l'optimisation globale. Ces méthodes intègrent des informations externes ou bien des connaissances a priori de la scène pour contraindre le problème.

Dans [ZK00], les auteurs proposent un algorithme coopératif pour calculer la carte de disparité en respectant les contraintes d'unicité et de continuité tout en détectant de manière explicite les parties cachées. Pour assurer la continuité de la surface, les valeurs de corrélation sont propagées dans un voisinage 3D. Une fois les valeurs stabilisées, les parties cachées sont détectées grâce à un seuillage sur les valeurs de corrélation. En effet, si une région est cachée, les scores de corrélation renvoyés sont faibles. On pénalise ainsi tous les pixels dont le score est faible.

Dans [AD04], les auteurs proposent un algorithme qui combine la corrélation dans des fenêtres locales de taille fixe et la minimisation d'énergie globale proposant ainsi deux disparités concurrentes dans une même fenêtre.

Un autre algorithme [LT02] alterne itérativement entre la segmentation de l'image en régions et l'approximation de chaque région par une surface de disparité lisse.

Dans [KSC01], on se place dans un contexte multi-vues. Les coûts d'appariement sont calculés localement, soit en utilisant des fenêtres adaptatives, soit en utilisant plusieurs fenêtres sur le même pixel. Ensuite, dans la formulation globale du problème, le terme d'attache aux données est la somme des scores de corrélation locaux. Le deuxième apport consiste à labelliser explicitement les pixels cachés dans l'une des images.

7.2.4 Bilan

On a proposé dans les sections précédentes un état de l'art détaillé des méthodes de minimisation d'énergie. Deux formulations se détachent : une formulation par programmation dynamique et une formulation à base de flots de graphes. Les méthodes coopératives sont basées sur l'une des deux formulations, mais intègrent également des connaissances a priori de la scène.

Le problème des méthodes basées sur la programmation dynamique est la difficulté de propager l'information inter-lignes épipolaires. Cet effet est visible sur le MNE utilisé en entrée de nos algorithmes, et qui est obtenu par programmation dynamique [Bai97]. La carte de disparité finale est calculée par bandes le long des épipolaires (cf. figure 6.10).

Pour les méthodes de minimisation d'énergie basées sur les flots de graphes, l'énergie à minimiser comporte deux termes : le terme d'attache aux données et le terme d'attache au modèle comprenant les connaissances a priori de la scène. Les améliorations les plus courantes dans la littérature portent sur la gestion des occultations et des parties cachées, la modélisation des discontinuités ou encore l'optimisation du temps de calcul puisque le problème est NP-complet. Pour gérer explicitement les occultations, Intille et Bobick [IB95] détectent les contours dans l'espace de disparité et diminuent le coût d'occultation le long des chemins pour ces contours, de telle façon qu'ils favorisent les discontinuités au niveau des gradients d'intensité. Dans [SSZ02], les auteurs utilisent une énergie de lissage dépendante de la différence d'intensité ce qui permet également de faire coïncider les discontinuités avec les changements d'intensité ou de couleur. La même idée a été utilisée dans [Fua93, BVZ99].

Cependant, la plupart de ces approches ne réalisent qu'une optimisation à partir d'une information pixellaire. Notre apport consistera à introduire de manière explicite des informations de plus haut niveau, composées de l'ensemble des primitives 3D extraites (cf. partie III), pour modéliser les discontinuités et régulariser les surfaces planes.

Le problème de reconstruction de surface hybride 3D devient alors un problème de minimisation globale contraint par des primitives 3D. La méthodologie d'optimisation 3D globale ainsi que la mise en œuvre des contraintes seront détaillées dans la section suivante.

7.3 Optimisation 3D contrainte

La modélisation globale de la scène doit prendre en compte les primitives extraites et tirer profit de leur complémentarité. Les primitives linéiques serviront à modéliser les discontinuités grâce à leur précision planimétrique. Les facettes 3D serviront à régulariser les pans de toits. Pour s'affranchir du problème de sous-détections des primitives, on s'autorise à utiliser les données

sources disponibles telles que les images et toute donnée directement et automatiquement obtenue à partir de celles-ci tel que le MNE. Pour mieux contraindre le problème, des données externes pourront également être utilisées, comme on le verra dans la section 7.3.6. En sortie du processus d'optimisation, on disposera d'une description 3D dense de la scène qui pourra être sous forme raster ou bien sous forme vectorielle.

7.3.1 Méthodologie

On se base sur la formulation du flot maximal proposée par [RC98] puisqu'elle permet d'introduire de manière naturelle les primitives 3D dans le processus d'optimisation. Dans [RC98], les auteurs montrent qu'une coupe minimale dans le cube de corrélation renvoie une carte de disparité. Étant donné un graphe \mathcal{G} , une source s et un puits t et une fonction de capacité de coupe notée $|C_{\mathcal{G}}|$, on a la relation suivante entre le flot maximal atteint sur l'ensemble des fonctions de flots valides et la coupe de capacité minimale.

$$Max(Flot(\mathcal{G})) = Min(|C_{\mathcal{G}}|) \quad (7.11)$$

Ce qui revient à dire qu'une recherche de flot maximal correspond à une coupe minimale dans le graphe 3D.

Cette formulation telle qu'elle est décrite par ses auteurs [RC98] ne modélise ni les occultations ni les discontinuités. Dans notre cas, les primitives 3D permettront de les gérer explicitement. Un autre inconvénient de leur formulation est de ne pouvoir modéliser que des pénalités linéaires pour les discontinuités. Cet inconvénient ne l'est pas dans un contexte de scènes urbaines denses où la plupart des discontinuités aux bords des bâtiments sont linéaires.

Compte tenu de ses caractéristiques, la formulation de Roy et Cox peut être facilement adaptée à notre contexte. L'optimisation sera réalisée à l'intérieur du cube de corrélation sur toute la scène 3D. Ce cube de corrélation est calculé une seule fois au début du processus et a été également utilisé pour qualifier les facettes 3D (cf. section 6.5.4.1). De plus, les scores de corrélation du cube étant calculés dans l'espace 3D, les images sont traitées de manière symétrique. En présence de primitives, les coûts des arêtes seront modifiés explicitement pour autoriser les discontinuités ou bien les pénaliser, privilégier une surface plane par rapport à une autre en fonction des facettes 3D extraites.

La figure 7.3 illustre le schéma global de l'optimisation hybride 3D proposée.

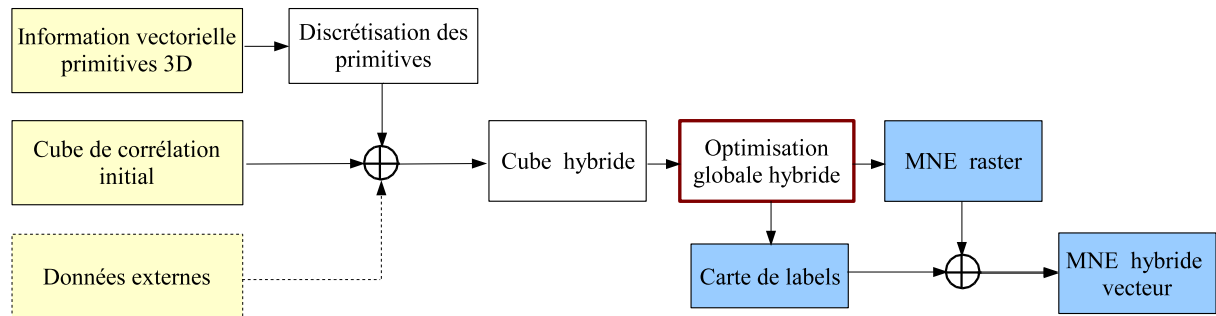


FIG. 7.3 – Schéma global de l'optimisation hybride

On dispose en entrée du cube de corrélation initial et de l'ensemble des primitives 3D extraites à partir des images (segments 3D, facettes 3D). On peut également utiliser en entrée de

l'algorithme des données externes telles que le réseau routier, des plans cadastraux ou bien un MNT pour modéliser le sol.

Une étape de discrétisation des primitives (cf. section 7.3.3.2) permettra d'injecter l'information vectorielle dans le cube. Cette étape permet d'utiliser les primitives 3D comme contraintes sans changer la structure du graphe initial. C'est essentiellement pour des raisons de commodité et par manque de temps, qu'on a choisi de discrétiser les primitives 3D. Il est envisagé dans les perspectives, de créer un graphe 3D hybride qui serait composé de voxels, mais également de nœuds correspondant à des primitives vectorielles.

Après l'étape de discrétisation des primitives, le cube devenu hybride sert à construire le graphe 3D et à réaliser l'optimisation globale afin d'obtenir un MNE hybride. Ce dernier sera sous forme raster puisque les primitives ont été discrétisées. La carte de labels générée comporte les différentes primitives qui ont été retenues dans la coupe. Cette carte de labels permet de remonter à l'information vectorielle correspondant aux primitives 3D. Le MNE final sera représenté en 3D, avec différents niveaux d'interprétation où, pour chaque région planimétrique, le niveau de primitive le plus fiable est retenu.

7.3.2 Construction du graphe 3D

Dans la formulation de flot maximal proposée par Roy et Cox [RC98], le graphe est construit de manière classique comme dans la plupart des problèmes de flot maximal : un maillage 3D de points (x, y, d) , où (x, y) sont les coordonnées images et d l'éventuelle disparité correspondante, la source et le puits. Le maillage est 6-connexe, la source est connectée au plan de disparité minimale et le puits est connecté au plan de disparité maximale.

Dans notre cas, le graphe sera construit à partir du cube de corrélation précalculé, le maillage 3D étant l'ensemble des voxels du cube. On se place dans l'espace objet, chaque voxel v est représenté par ses coordonnées 3D (x, y, z) et son score de corrélation $Corr(v)$ et on rajoute la source et le puits. On obtient ainsi un graphe pondéré $\mathcal{G} = \langle \nu, \mathcal{E} \rangle$, ν étant l'ensemble des nœuds et \mathcal{E} l'ensemble des arêtes non orientées $\mathcal{E} \in \nu^2$, avec deux nœuds terminaux s et t , la source et le puits. La source est connectée, pour chaque colonne, aux voxels d'altitude minimale et le puits aux voxels d'altitude maximale (cf. figure 7.4(a)). Le maillage est 6-connexe (cf. figure 7.4(b)).

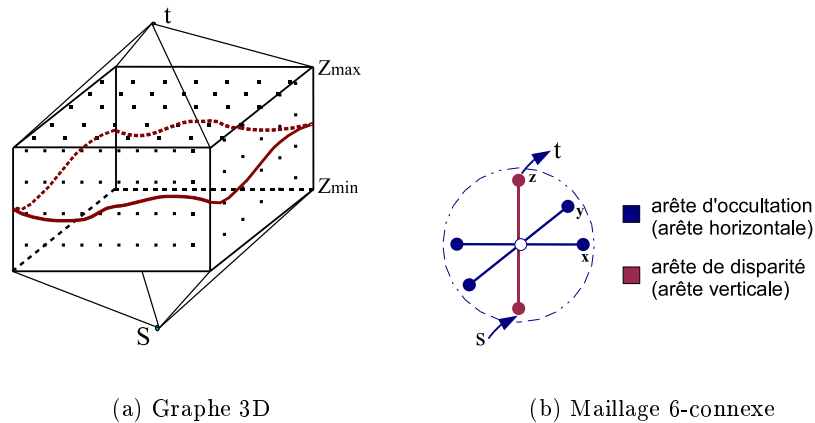


FIG. 7.4 – Construction du graphe 3D

Le graphe est non orienté et $\forall (u, v) \in \mathcal{G}$, la capacité $C(u, v) \geq 0$.

Afin de contrôler la régularité de la surface obtenue, il est indispensable de distinguer deux types d'arêtes. Les arêtes suivant l'axe z auront des capacités C_{disp} , la disparité étant liée à l'altitude. Toutes les autres arêtes auront des capacités C_{occ} . On verra dans la section 7.3.7 que la capacité d'occultation contrôle le degré de lissage de la surface finale.

$$C(u, v) = \begin{cases} 0 & \text{si } (u, v) \notin \mathcal{G} \\ \infty & \text{si } u = s \text{ ou } v = t \\ C_{disp} & \text{si } (u - v) = (0, 0, \Delta z) \\ C_{occ} & \text{si } (u - v) = (\Delta x, 0, 0) \text{ ou } (u - v) = (0, \Delta y, 0) \end{cases} \quad (7.12)$$

Les capacités des arêtes s'expriment en fonction des scores de corrélation. Plus le score de corrélation est fort, plus les voxels sont fiables et plus la capacité est faible. On définit le coût d'un nœud u comme suit :

$$\text{Coût}(u) = 100 * (1 - \text{Corr}(u)) \quad (7.13)$$

avec $\text{Corr}(u)$ le coefficient de corrélation normalisé centré de u .

$$\begin{cases} C_{disp}(u, v) = \frac{\text{Coût}(u) + \text{Coût}(v)}{2} \\ C_{occ}(u, v) = k * C_{disp}(u, v) + C_f \end{cases} \quad (7.14)$$

- u et v deux nœuds adjacents du graphe
- $C_{disp}(u, v)$ étant la capacité d'une arête verticale (capacité de disparité)
- $C_{occ}(u, v)$ étant la capacité d'une arête horizontale (capacité d'occultation)
- k étant un coefficient de lissage et C_f un coût fixe

On rappelle que la capacité d'une coupe de graphe $|\mathcal{C}_G|$ est égale à la somme des capacités des arêtes coupées et qui séparent ν^s et ν^t .

$$|\mathcal{C}_G| = \sum_{(u,v) \in (\nu^s, \nu^t)} |C(u, v)| \quad (7.15)$$

Pour des chemins de coupe verticaux, le coût global de la coupe est exprimé en fonction des capacités d'occultation, alors que les chemins de coupe horizontaux font intervenir les capacités C_{disp} (cf. figure 7.5).

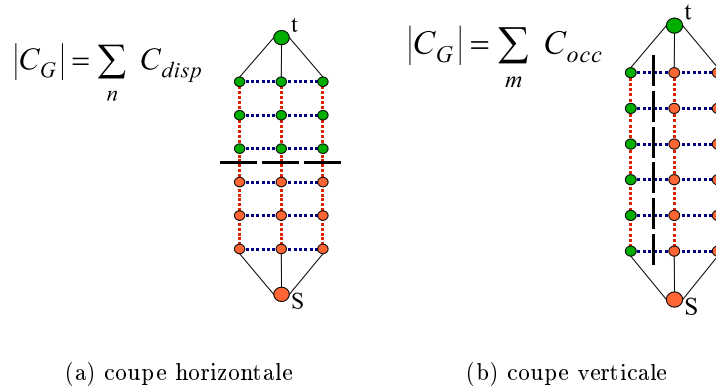


FIG. 7.5 – Capacités de coupe de graphe

Le paramètre k de la capacité d'occultation (cf. Eq 7.14) permet de régler le lissage de la surface finale. Plus k est grand, plus les discontinuités sont pénalisées. En effet, si k augmente, la capacité C_{occ} augmente, on favorise ainsi la coupe par les arêtes verticales dont le coût est moins élevé. Les discontinuités sont pénalisées et la surface finale est plus régulière. Le coût fixe C_f permet de pénaliser d'un même score minimal toutes les discontinuités quelle que soit la capacité des arêtes correspondantes. Ceci permettra de mieux gérer les discontinuités effectives modélisées par la présence de segments 3D (cf. section 7.3.5).

7.3.3 Gestion du graphe 3D hybride

7.3.3.1 Utilisation du cube de corrélation initial

L'information vectorielle extraite à partir des images (segments 3D, facettes 3D, sol) n'est pas dense comme on a pu le voir dans la partie III. D'une part, cette information n'est pas dense en 2D, il en est ainsi de même en 3D. D'autre part, à cause de la faible précision de reconstruction altimétrique, des primitives jointives en 2D ne le sont plus forcément en 3D. Par conséquent, afin d'obtenir un MNE hybride final qui soit **dense**, il est indispensable de fournir l'information manquante lors de l'optimisation globale.

Pour cela, le graphe 3D sera construit à partir du cube de corrélation initial et les primitives 3D seront injectées ponctuellement et de manière explicite dans le graphe. L'information vectorielle sera ainsi exploitée, si elle est disponible. Sinon, l'information pixellaire provenant du cube de corrélation sera utilisée. On obtiendra, après optimisation, un MNE hybride fournissant plusieurs niveaux de description de la scène 3D.

7.3.3.2 Discrétisation des primitives

Afin de garder une structure de graphe simple permettant d'utiliser tout type de données externes, on a choisi de discrétiser les primitives 3D pour les intégrer dans le cube de corrélation. L'utilisation de toute information externe devient ainsi très simple en respectant les étapes suivantes :

- discrétisation des données vectorielles
- projection des primitives dans le cube
- construction du graphe 3D à partir des voxels du cube
- gestion explicite des nœuds correspondants aux primitives 3D

La recherche de la coupe minimale dans le graphe 3D renvoie une surface $2D^{1/2}$ où chaque point (x, y) de la scène est affecté de la valeur d'altitude du voxel correspondant retenu lors de la coupe. On obtient ainsi, un MNE hybride sous forme rasterisée. Grâce à la génération d'une carte de labels des différentes primitives retenues dans la coupe, on peut remonter à l'information vectorielle correspondante à chaque voxel et obtenir ainsi une représentation 3D du MNE hybride.

7.3.3.3 Gestion du graphe 3D

7.3.3.3.1 Implémentation par nappes La projection des primitives vectorielles dans le cube apporte des informations supplémentaires sur les altitudes et permet de mieux focaliser sur les bâtiments. Elle permet également de diminuer le nombre de nœuds dans le graphe. En effet, au lieu de garder une structure de cube rigide, on propose une implémentation par nappes permettant de focaliser sur les bâtiments. On détermine deux nappes d'altitude minimale et maximale entre lesquelles les scores de corrélation sont élevés ($Corr(u) > 50\%$) ou bien au-delà desquelles les nœuds sont interdits, suite à l'utilisation de facettes 3D (cf. section 7.3.4.4). Le

graphe 3D sera construit à partir des nœuds situés entre les deux nappes. La source s est reliée à la nappe d'altitude minimale et le puits t à la nappe d'altitude maximale (cf. figure 7.6). Cette implémentation par nappes permet un gain de temps considérable lors de l'optimisation globale.

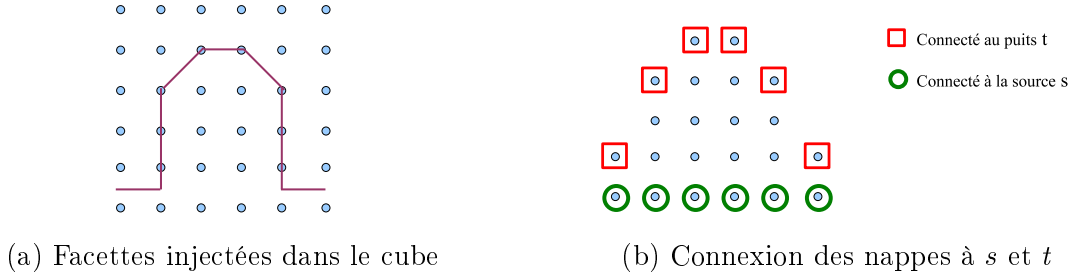


FIG. 7.6 – Connexion des nappes dans le graphe \mathcal{G}

On développera dans ce qui suit, les différentes configurations rencontrées au niveau de la gestion des nappes.

7.3.3.3.2 Validité du graphe 3D D'après [RC98], pour qu'une coupe de capacité minimale dans le graphe 3D renvoie une carte de disparité, il faut qu'il y ait un chemin de capacité finie entre la source et le puits. Pour une colonne du cube (x, y, z) $z \in [z_{min}, z_{max}]$, la source s étant connectée au nœud d'altitude minimale et le puits t au nœud d'altitude maximale, $s(x, y, z_{min})$ et $t(x, y, z_{max})$ sont des arêtes de capacité infinie (eq. 7.12). Pour assurer la validité du graphe 3D, il faut qu'il y ait au moins une arête de capacité finie entre deux nœuds d'une même colonne.

7.3.3.3.3 Conditions pour assurer la contrainte d'unicité La contrainte d'unicité revient à assurer que le chemin de capacité minimale ne coupe qu'une et une seule arête de coordonnées (x_i, y_i) . Autrement dit, elle permet d'éviter les surplombs lors de la coupe finale. Le problème de surplombs est un problème délicat en stéréovision surtout en visée quasi-verticale. Ce problème n'a pas été traité dans [RC98] puisque chaque point (x_i, y_i) est affecté de la plus grande disparité associée à l'arête correspondante, retenue dans la coupe.

Dans notre cas, pour assurer la contrainte d'unicité, on impose une condition suffisante sur les capacités des arêtes horizontales autorisées : $C_{Ha} = C_{ste}$.

Le problème des surplombs est illustré figure 7.7 en 2D dans le cas d'utilisation de facettes 3D. La condition est assez intuitive. Les facettes F_1, F_2 et F_3 sont représentées en trait noir continu. Les chemins de coupes verticaux sont illustrés en rouge et les chemins horizontaux en vert.

En effet, on voit sur la figure 7.7, que les deux chemins proposés coupent le même nombre d'arêtes horizontales. Si $C_{Ha} = C_{ste}$, les chemins verticaux ont le même coût dans les deux cas de figures. Par conséquent, la coupe qui minimisera la capacité totale est celle qui minimise le nombre d'arêtes verticales coupées, autrement dit celle qui évite les surplombs. Dans l'exemple illustré ci-dessus, la coupe A sera retenue.

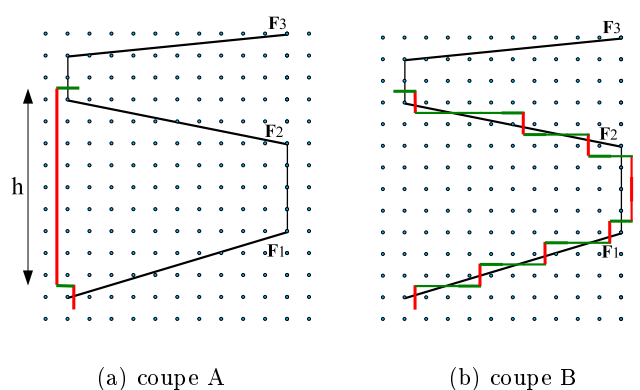


FIG. 7.7 – Problème de surplombs

7.3.4 Régularisation par surfaces planes : facettes 3D

Les facettes 3D constituent des primitives plus fiables que les primitives linéiques, d'une part parce que leur taille est plus grande que les fenêtres de corrélation ce qui rend les scores plus fiables que dans le cas des segments 3D. D'autre part, avec un faible rapport B/H , la précision altimétrique des plans 3D est meilleure puisque leur estimation est réalisée à partir d'un plus grand nombre de points 3D. En outre, les facettes correspondent directement aux toits de bâtiments et fournissent ainsi deux informations importantes : la régularité du toit et sa délimitation. Dans la section suivante, l'utilisation des facettes 3D sera détaillée.

7.3.4.1 Principe

La figure 7.8 schématise un résultat idéal qu'on aurait avec le processus d'optimisation globale en utilisant les facettes 3D fiables. Une coupe en 2D du cube de corrélation est présentée.

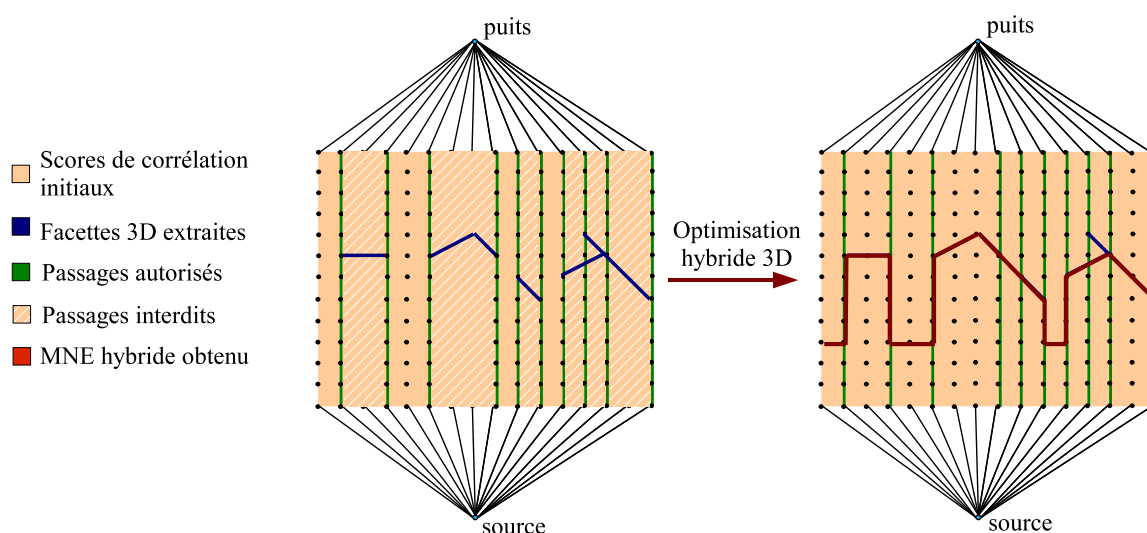


FIG. 7.8 – Exemple idéal d'optimisation hybride avec les facettes 3D

Le principe est le suivant : Si on dispose d'information vectorielle, on l'exploite en forçant

le flot à passer par les facettes détectées. Pour cela, on interdit le passage du flot au-dessus et en-dessous des facettes 3D (cf. section 7.3.4.5) et on l'autorise à passer sur les contours des facettes 3D en créant des "couloirs de passage " (cf. eq 7.16). Ces derniers permettent ainsi au flot d'atteindre toutes les facettes.

L'optimisation hybride devra gérer deux problèmes liés aux détecteurs de facettes 3D. Pour résoudre le problème de sous-détection des facettes 3D, l'algorithme devra utiliser les scores de corrélation initiaux en assurant la continuité entre les facettes 3D. Le deuxième problème est la superposition de facettes, qui peut être due à des problèmes de délimitations de facettes ou bien à plusieurs hypothèses de facettes concurrentes pour une même région planimétrique. Les capacités d'arêtes doivent être choisies de telle sorte qu'elles pénalisent les passages intempestifs entre facettes superposées (cf. section 7.3.4.5).

Pour utiliser les facettes 3D dans le processus d'optimisation globale, trois étapes importantes seront détaillées dans ce qui suit. Tout d'abord, l'étape de discrétisation de la facette où il faudra veiller à ce que le graphe 3D reste valide, ensuite l'étape de gestion des nœuds pour l'interdiction de voxels et la création de couloirs de passage et enfin la gestion des nappes correspondantes aux facettes afin de préserver les discontinuités.

7.3.4.2 Discrétisation des facettes 3D

Les facettes sont discrétisées suivant un maillage en 6-connexité (cf. section 7.3.2).

Pour que la coupe puisse passer par une facette donnée, il faut garantir la validité du graphe 3D (cf. section 7.3.3.2). Pour cela, il faut que pour tout point (x, y) , il existe au moins une arête de la forme $(x, y, z + 1) - (x, y, z)$. En termes de voxels, cette condition se traduit par l'existence d'au moins deux voxels consécutifs sur une même colonne. Par exemple, pour une facette horizontale $z = Cste$, les voxels doivent être dédoublés à l'altitude $z + 1$ pour permettre à la coupe de passer par cette facette. Enfin, les voxels correspondants sont affectés du score de corrélation moyen de la facette (cf. section 6.5.4.2).

7.3.4.3 Création des couloirs de passage

Pour permettre le passage entre différentes facettes, on crée des "couloirs de passage" sur les bords de facettes (cf. figure 7.9) qui sont définis comme suit :

$$C_F = \bigcup_i^n \bigcup_{k=z_{min}^i}^{z_{max}^i} (x_i, y_i, k) / (x_i, y_i) \in \oplus \text{Cont}(F) \quad (7.16)$$

n est le nombre de points du contour dilaté de la facette F .

z_{min}^i et z_{max}^i sont les altitudes minimale et maximale du pixel (x_i, y_i) .

Les contours des facettes sont dilatées de 1 pixel afin de créer des arêtes horizontales de très faibles coûts qui favorisent le passage de la coupe par ces couloirs. La capacité des arêtes au niveau des couloirs de passage est choisie faible mais non nulle pour pouvoir prendre en compte la magnitude des discontinuités. L'inconvénient apparaît dans le cas de plusieurs hypothèses de facettes concurrentes où les facettes supérieures sont forcément plus pénalisées à cause de la magnitude de la discontinuité.

7.3.4.4 Préservation des discontinuités

Sans l'utilisation des primitives 3D, la gestion des nappes se fait en connectant, pour chaque colonne, les nœuds d'altitude minimale et maximale respectivement à la source et au puits (cf. section 7.3.3.3.1). Dans ce cas, on est assurés d'avoir des nappes jointives. Avec l'utilisation des facettes 3D et l'interdiction de certains nœuds, les nappes ne sont plus forcément jointives, il faut donc gérer de manière explicite les coûts des arêtes et veiller à la validité du graphe 3D pour pouvoir passer d'une nappe à une autre.

Pour la gestion des nappes, il est important de repositionner l'étude dans le contexte urbain dense. Le paysage urbain est caractérisé par de fortes discontinuités le long des façades. La coupe doit préserver les discontinuités. Grâce à l'optimisation hybride, les discontinuités traitées vont surtout correspondre à des couloirs de passage entre les différentes facettes. Il faut être capable d'estimer la magnitude des discontinuités soit pour les pénaliser, soit pour leur donner plus de poids. Différents cas peuvent se présenter : soit, on est dans le cas d'une seule hypothèse de facette par région planimétrique et la magnitude des discontinuités n'aura aucun effet sur la coupe puisqu'elle sera forcée à passer par cette facette. Soit, on est dans le cas de plusieurs facettes concurrentes et il faudra prendre en compte la magnitude des discontinuités pour choisir l'une ou l'autre des facettes après le couloir de passage. On aura également les cas de recouvrement planimétrique partiel ou bien d'intersection entre des facettes concurrentes, la connaissance de la magnitude des discontinuités permettra dans ce cas d'éviter les passages intempestifs entre ces facettes.

En pratique, pour prendre en compte la magnitude des discontinuités lors de la coupe, on propose de reconnecter tous les nœuds se trouvant au bord des nappes min et max respectivement à la source et au puits et pas seulement les nœuds d'altitude minimale et maximale de chaque colonne. La figure 7.9 illustre la gestion de la nappe dans le cas d'une transition entre deux facettes avec la présence des couloirs de passage sur les bords de facettes.

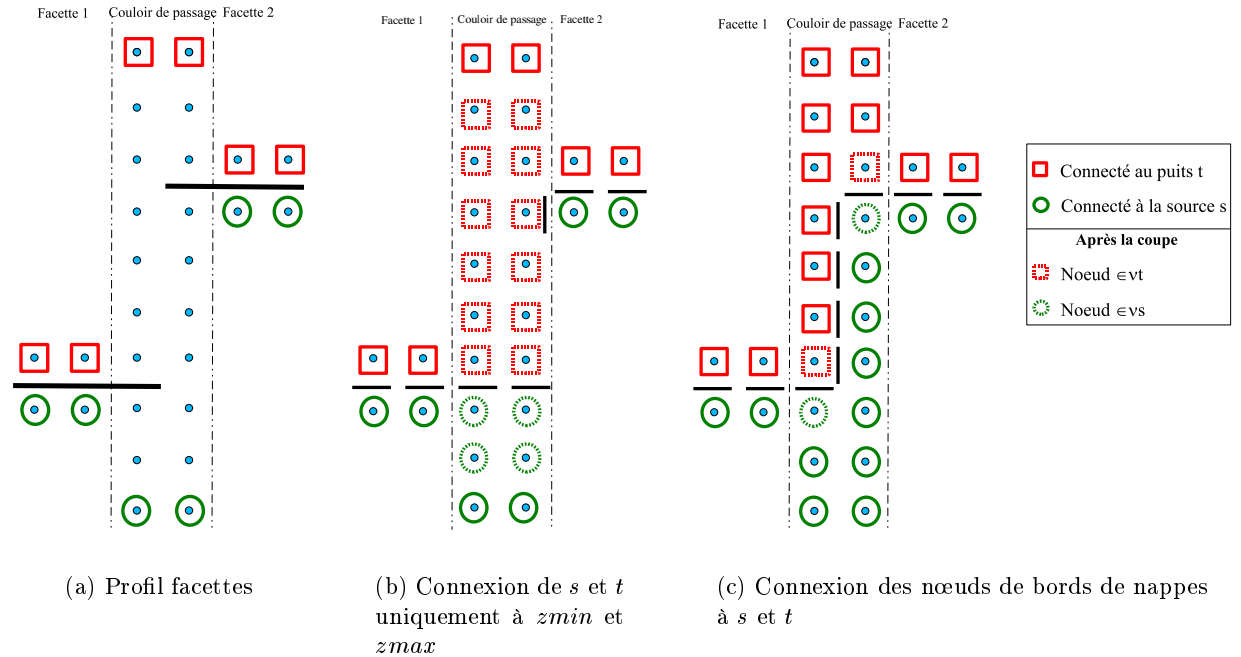


FIG. 7.9 – Gestion des nappes en présence de facettes dans le graphe \mathcal{G}

Les chemins de coupe sont illustrés en trait discontinu gras. Suite à l'optimisation globale, la coupe minimale divise le graphe G en deux ensembles ν^s et ν^t reliés respectivement à la source et au puits. L'appartenance des nœuds à ces deux ensembles après la coupe est également illustrée.

La figure 7.9(b) montre le cas où la source et le puits sont uniquement reliés aux voxels d'altitude minimale et maximale de chaque colonne. Après la coupe, tous les voxels du couloir de passage se retrouvent dans le même ensemble et par conséquent la coupe ne passe pas effectivement par ces discontinuités qui ne sont pas prises en compte dans le calcul du coût de la coupe.

La figure 7.9(c) montre le cas où la source et le puits sont connectés aux nœuds sur le bord des nappes. Les couloirs de passage sont alors reliés de part et d'autre à s et t . Par conséquent, la coupe passera effectivement dans les couloirs de passage et toutes les discontinuités seront prises en compte. Le fait de reconnecter tous les nœuds de bord des nappes min et max implique une dilatation des couloirs de passage pour garantir un chemin de capacité finie entre s et t (cf. figure 7.9). Ceci conforte la dilatation des contours de facettes pour créer des arêtes horizontales de très faibles coûts qui favorisent le passage de la coupe par ces couloirs.

7.3.4.5 Capacités des arêtes en présence de facettes 3D

7.3.4.5.1 Gestion des arêtes interdites Pour forcer le passage de la coupe par les facettes 3D injectées dans le cube, les nœuds qui sont au-dessus et en-dessous d'une facette donnée sont interdits. L'interdiction d'un nœud est réalisée en affectant aux arêtes verticales correspondantes (arêtes Verticales Non Autorisées) une capacité infinie. En pratique, $C_{Vna} = 100$.

7.3.4.5.2 Gestion des facettes concurrentes Dans le cas de superposition de facettes concurrentes pour une même région planimétrique, les nœuds se situant entre les différentes facettes sont interdits (cf. figure 7.8) c.à.d que les arêtes horizontales et verticales correspondantes ne sont pas autorisées et sont par conséquent affectées respectivement de C_{Hna} et C_{Vna} . Les capacités d'arêtes doivent être choisies de façon à éviter le passage intempestif entre facettes superposées. Ce choix sera détaillé dans ce qui suit.

La figure 7.10 illustre sur un exemple simple, la configuration de deux facettes superposées ainsi que les chemins de coupes possibles.

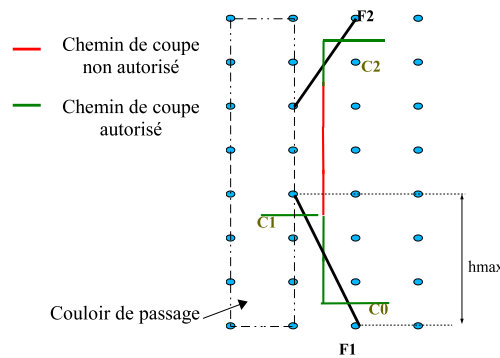


FIG. 7.10 – Cas de facettes superposées - Illustration des chemins de coupes en partant de C_1

Problème : Si dans le couloir de passage, la coupe passe par F_1 , quelles capacités faut-il choisir pour éviter la facette F_2 ?

Pour résoudre ce problème, il faut connaître le nombre maximal d'arêtes verticales et horizontales pouvant être coupées afin d'estimer le coût de passage d'une facette à une autre. Soit α_{max} ,

la pente maximale des facettes de toits, le nombre maximal d'arêtes horizontales coupées n_{hmax} , et d'arêtes verticales coupées n_{xmax} sur une facette sont calculés comme montré sur la figure 7.10. Δ_{xy} et Δ_z sont les résolutions en xy et en z du cube de corrélation. L étant la longueur de la facette.

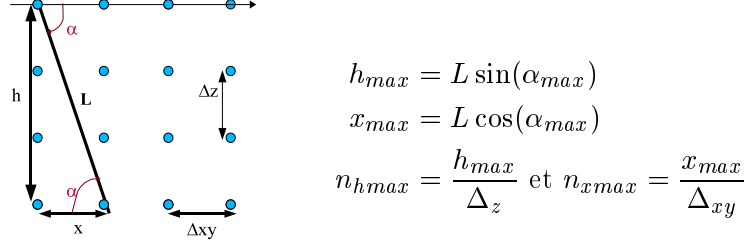


FIG. 7.11 – Illustration du nombre d'arêtes coupées pour une facette

On note $|C_F|$ la capacité de la coupe qui passe par la facette F . Sur la figure 7.10, C_0 , C_1 et C_2 représentent les capacités des chemins horizontaux, donc celles des arêtes verticales autorisées respectivement sur les facettes F_1 et F_2 . C_{Hna} et C_{Ha} sont les capacités des arêtes horizontales respectivement non autorisées et autorisées, coupées par le chemin.

$$\begin{aligned}
 |C_{F_1}| < |C_{F_2}| &\Leftrightarrow C_1 + \sum C_{V_a}(F_1) + \sum C_{H_a}(F_1) < C_1 + \sum C_{Hna} + \sum C_{V_a}(F_2) \\
 &\Leftrightarrow \sum C_{Hna} > max(\sum C_{V_a}(F_1) - \sum C_{V_a}(F_2)) + max(\sum C_{H_a}) \\
 &\Leftrightarrow \sum C_{Hna} > n_{xmax} * 100 + n_{hmax} * C_{Ha} \\
 &\Leftrightarrow \sum C_{Hna} > L \left(\frac{\cos(\alpha_{max})}{\Delta_{xy}} * 100 + \frac{\sin(\alpha_{max})}{\Delta_z} * C_{Ha} \right)
 \end{aligned} \tag{7.17}$$

Si on se replace dans le graphe 3D, il faut considérer la surface maximale de la facette pour la capacité totale :

$$C_{Hna} > S_{max} \left(\frac{\cos(\alpha_{max})}{\Delta_{xy}} * 100 + \frac{\sin(\alpha_{max})}{\Delta_z} * C_{Ha} \right) \tag{7.18}$$

Le tableau 7.1 résume les différentes capacités affectées aux arêtes horizontales et verticales suivant qu'elles soient autorisées ou pas.

TAB. 7.1 – Capacités des arêtes du graphe

	Autorisées	Non autorisées
C_V	$C_{V_a} = C_{disp}$	$C_{Vna} = 100$
C_H	$C_{H_a} = C_{ste}$	$C_{Hna} > S_{max} \left(\frac{\cos(\alpha_{max})}{\Delta_{xy}} * 100 + \frac{\sin(\alpha_{max})}{\Delta_z} * C_{Ha} \right)$

7.3.5 Modélisation des discontinuités : segments 3D

Grâce au faible rapport B/H du couple stéréoscopique, les segments 3D extraits sont bien positionnés en planimétrie, par contre ils souffrent d'une faible précision altimétique. Par conséquent, ce sont essentiellement les coordonnées planimétriques qui seront utilisées pour modéliser les discontinuités. Dans notre cas, la fiabilité planimétrique des segments 3D est assurée par l'élagage suivant leur orientation par rapport à la base(cf. chapitre 4). Si on envisage l'utilisation des

segments 3D dans un autre contexte (images aériennes, grand $B/H..$) et avec d'autres extracteurs de primitives linéiques, il faut s'assurer de la bonne précision planimétrique des segments qui est une condition nécessaire pour la modélisation des discontinuités.

Les discontinuités pourront être soit de forte magnitude (façades de bâtiments), soit de faibles magnitudes (faîtes de toits). Pour mettre à profit la précision planimétrique, des "couloirs de passage" sont créés le long des segments 3D pour autoriser explicitement les discontinuités.

Le couloir de passage correspondant à un segment S donné est défini comme suit :

$$C_S = \bigcup_{i=0}^n \bigcup_{k=z_{min}^i}^{z_{max}^i} (x_i, y_i, k) / (x_i, y_i) \in \oplus P_s \quad (7.19)$$

avec P_S l'ensemble des pixels de S .

z_{min}^i et z_{max}^i sont les altitudes minimale et maximale du pixel (x_i, y_i) .

De la même façon que pour les facettes 3D (cf. section 7.3.4.3), les couloirs de passages relatifs aux segments sont dilatés pour créer des arêtes horizontales à faibles coûts favorisant le passage par ces couloirs.

7.3.5.0.3 Capacités des arêtes en présence d'un segment On cherche à favoriser les discontinuités le long des couloirs de passage, les arêtes correspondantes sont affectées de la valeur minimale d'une capacité soit $C_{disp} = 0$. On rappelle ici l'expression de la capacité d'occultation $C_{occ}(u, v) = k * C_{disp}(u, v) + C_f$ (cf. eq 7.14).

Pour mettre en évidence l'apport des segments 3D, la constante de lissage k doit être différente de 0. En effet,

- Si $k = 0$, toutes les arêtes horizontales ont un coût fixe, par conséquent toutes les discontinuités sont pénalisées de la même façon ($C_{occ}(u, v) = C_f$) et l'utilisation des segments n'aura aucun effet dans ce cas. Le résultat de l'optimisation avec ou sans segments est exactement le même (cf. figure 7.12-a).
- Si $k \neq 0$ et $C_{disp} = 0$, la capacité d'occultation sera à son minimum ($C_{occ}(u, v) = C_f, (u, v) \in C_s$) dans les couloirs de passage reliés aux segments. On favorise ainsi le passage de la coupe par ces couloirs de passage (cf. figure 7.12-b-c). Les bords de bâtiments sont mieux délimités que dans le cas de l'optimisation sans primitives où le paramètre k a pour effet de lisser énormément la surface finale.

Si $C_f = 0$, le coût de passage par les discontinuités devient nul et leur magnitude n'est plus prise en compte lors de la coupe. Ce problème est bien illustré sur la figure 7.12-b, où l'altitude maximale est retenue lors du passage par les segments.

Si $C_f \neq 0$, la magnitude des discontinuités est prise en compte et l'altitude au niveau des segments est mieux estimée en tenant compte de la régularité de la surface (cf. figure ??-b).

Les segments 3D peuvent être également utilisés avec les conditions suivantes : $k \neq 0$ et $C_{disp} \neq 0$ pour les couloirs de passage correspondants. Cependant, dans ce cas, la condition (1) n'est plus vérifiée dans tout le graphe et la contrainte d'unicité n'est assurée qu'en présence de primitives. Ailleurs, l'altitude maximale sera retenue.

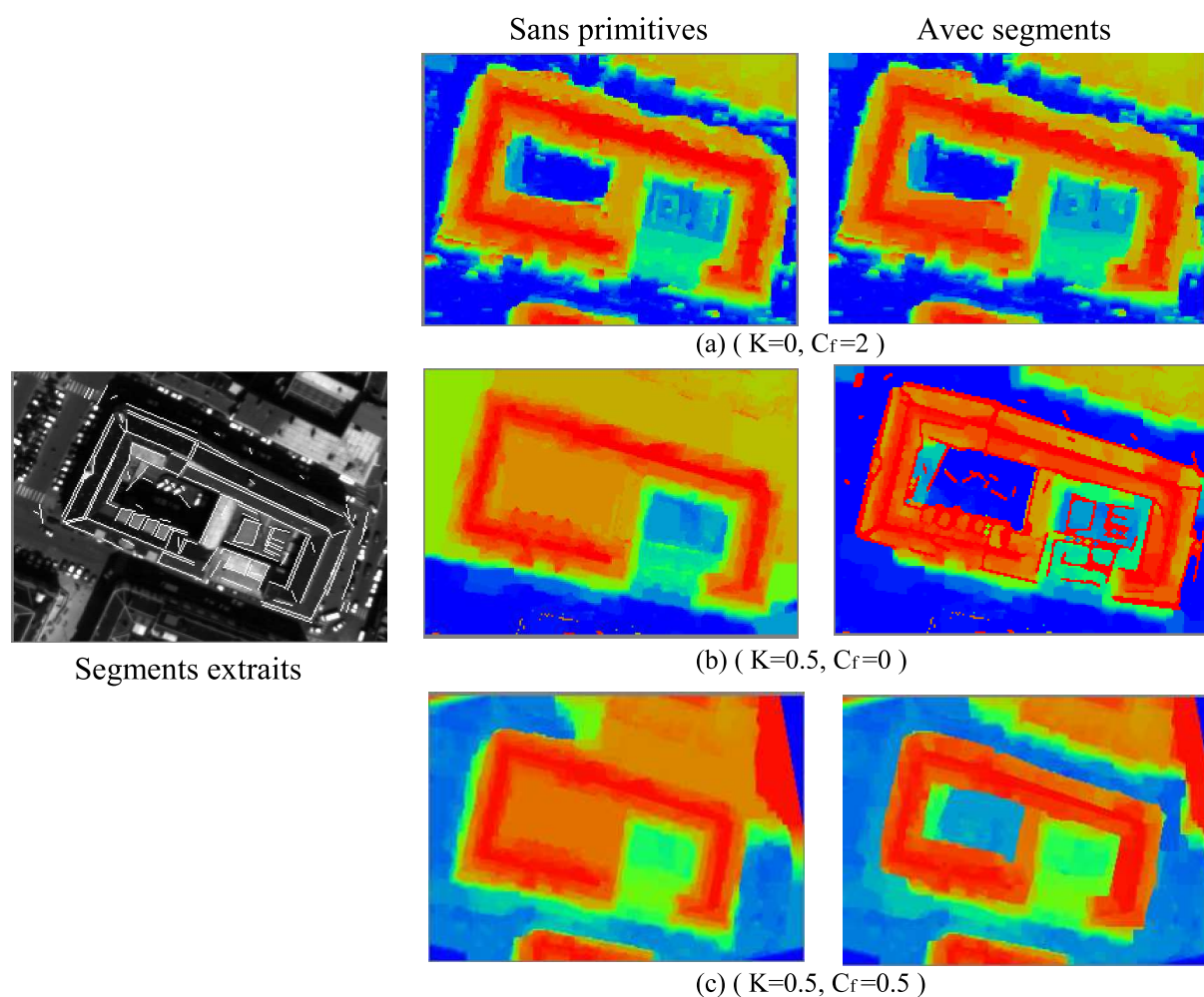


FIG. 7.12 – Optimisation globale avec les segments 3D

7.3.6 Modélisation du sol

Afin de pallier aux problèmes de facettes non fiables ou de débordements au niveau des bords de toits, on propose d'utiliser le sol dans le processus d'optimisation globale. Ceci nécessite une première étape d'extraction du sol et une deuxième étape de modélisation. Dans le cadre de cette thèse, on ne s'est pas intéressé à l'extraction du MNT (Modèle Numérique de Terrain). On utilise une classification semi-automatique basée sur une croissance de régions à partir de germes choisis par un opérateur. Cette étape est réalisée suite à la modélisation des discontinuités par les segments 3D qui permet de régulariser les bords de bâtiments et donc d'avoir une meilleure estimation du sol. Dans les perspectives, on peut également envisager d'utiliser des données externes (réseau routier 3D ou limites cadastrales) pour l'extraction du sol. Dans un deuxième temps, le sol est modélisé comme une facette qui serait une nappe minimale sur toute la scène. Elle sera injectée dans le cube de corrélation et sera mise en concurrence avec toute autre facette de toit. Les capacités d'arêtes relatives à la nappe du sol seront gérées de la même façon que pour les facettes 3D. Par contre, il n'y aura pas de création de couloirs sur les bords puisque la nappe du sol couvre toute la scène.

7.3.7 Résultats

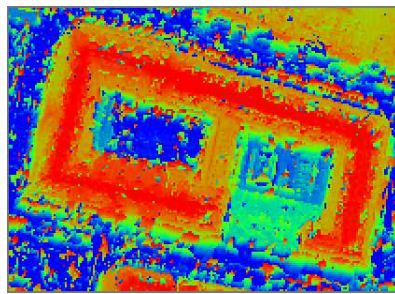
7.3.7.1 Optimisation sans primitives

La figure 7.13 illustre des résultats d'optimisation sans primitives. La figure 7.13(a) montre l'influence du coût fixe C_f .

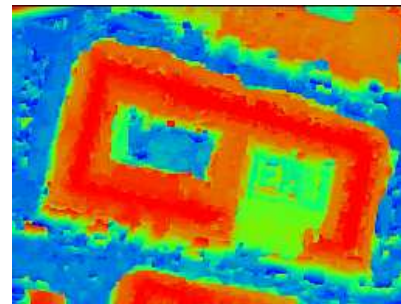
Si ($K = 0, C_f = 0$), seule l'énergie d'attache aux données est prise en compte. Le résultat obtenu est le même qu'en appliquant "The Winner Takes All", c.à.d que pour chaque pixel (x, y) , l'altitude qui renvoie le maximum du score de corrélation est retenue. Le MNE final est très bruité.

Si ($K = 0, C_f \neq 0$), les discontinuités sont pénalisées d'un même score minimal et leur magnitude est prise en compte lors de la coupe. Le MNE obtenu est plus régulier.

La figure 7.13(b) illustre l'effet du paramètre de lissage k . Plus k augmente, plus la surface finale est lissée et les discontinuités pénalisées.

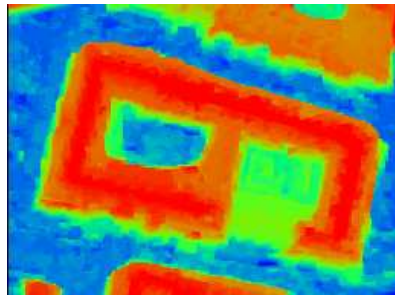


($K = 0, C_f = 0$)

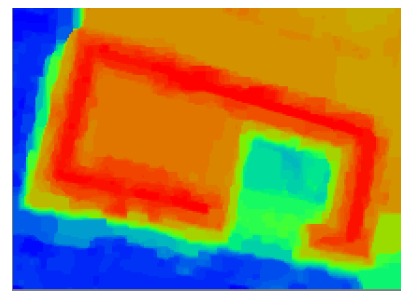


($K = 0, C_f = 0.5$)

(a) Influence de la constante C_f



($K = 0.1, C_f = 0.5$)



($K = 0.5, C_f = 0.5$)

(b) Influence du paramètre de lissage k

FIG. 7.13 – Optimisation sans primitives

7.3.7.2 Optimisation avec primitives

Des résultats d'optimisation avec primitives sont présentés sur des images aériennes de la ville d'Amiens à 50 cm, $B/H = 0.2$. Le chapitre 9 présente plusieurs résultats et évaluations sur différentes zones de test et avec différents paramétrages.

La figure 7.14 montre un résultat d'optimisation sur une zone d'Amiens à 50 cm en utilisant les segments 3D. Les bords de bâtiments sont nettement mieux délimités grâce à la présence des segments 3D. Cet effet est très visible au niveau de la cour intérieure du bâtiment en "U". L'algorithme est toutefois sensible à l'exhaustivité des segments 3D. En absence de segments 3D, le résultat obtenu est le même que celui de l'optimisation sans primitives (cas des routes entre bâtiments).

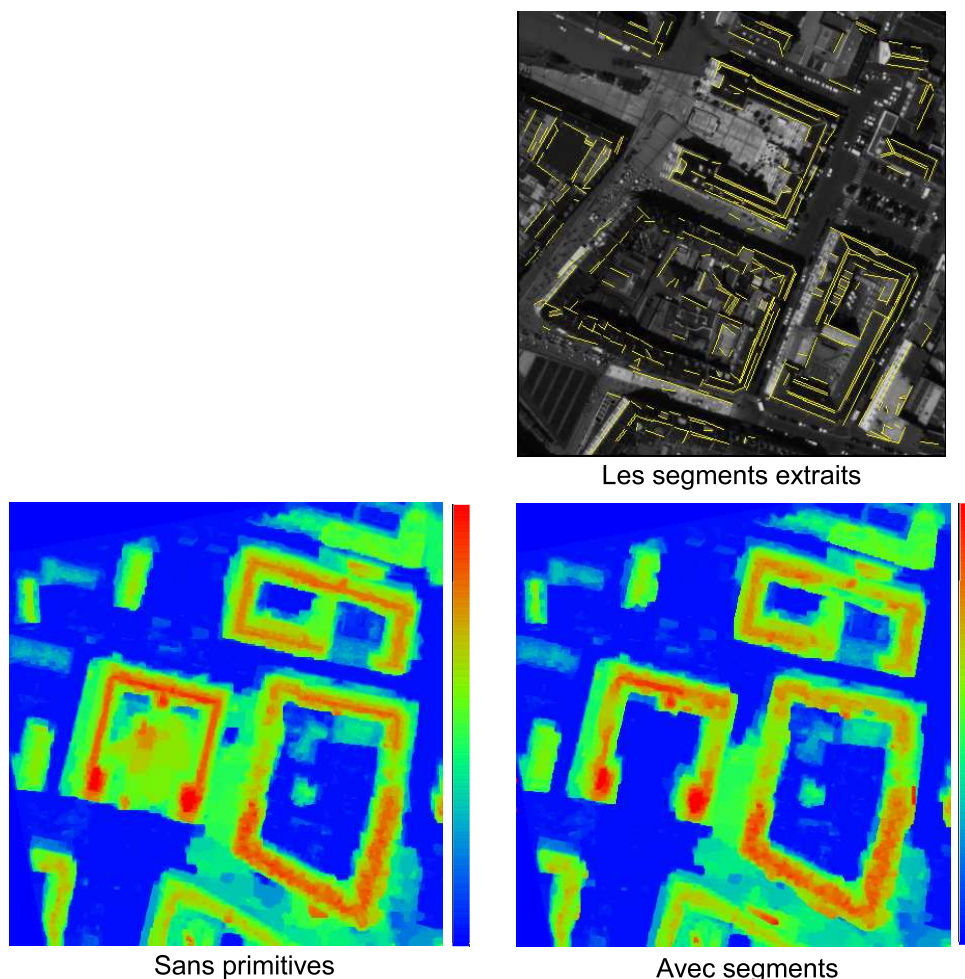
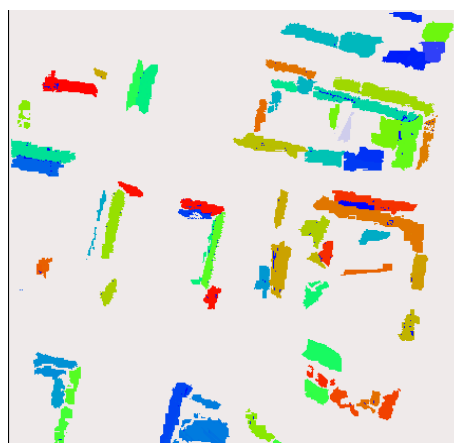


FIG. 7.14 – Exemple d'optimisation avec les segments ($K = 0.2$, $C_f = 0.5$)

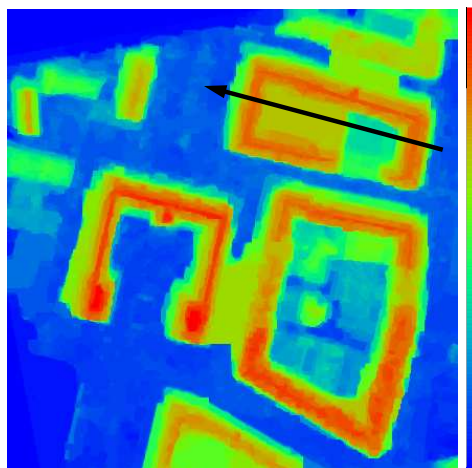
La figure 7.15 montre un résultat d'optimisation sur la même zone avec les facettes 3D. L'utilisation des facettes permet de régulariser le MNE final au niveau des toits.

Pour mettre en évidence cette régularisation, des profils sur le MNE permettent de comparer les résultats avec et sans facettes. On remarque que les pentes de toits sont mieux modélisées en présence de facettes et les bords de bâtiments sont également mieux délimités. Cet effet est visible au niveau de la cour intérieure du bâtiment "9".

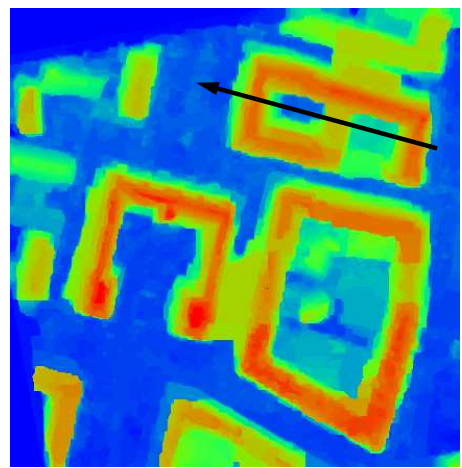
Par contre, on peut relever deux problèmes ; tout d'abord un problème de discrétisation des facettes qui est visible au niveau des pentes de bâtiments. Cet effet est relié au pas de discrétisation du cube de corrélation. Si on affine la discrétisation, on augmente le nombre de nœuds du graphe 3D et par conséquent la combinatoire. Un compromis est à trouver entre le pas de discrétisation et le temps de calcul. D'autre part, on remarque un problème de débordements



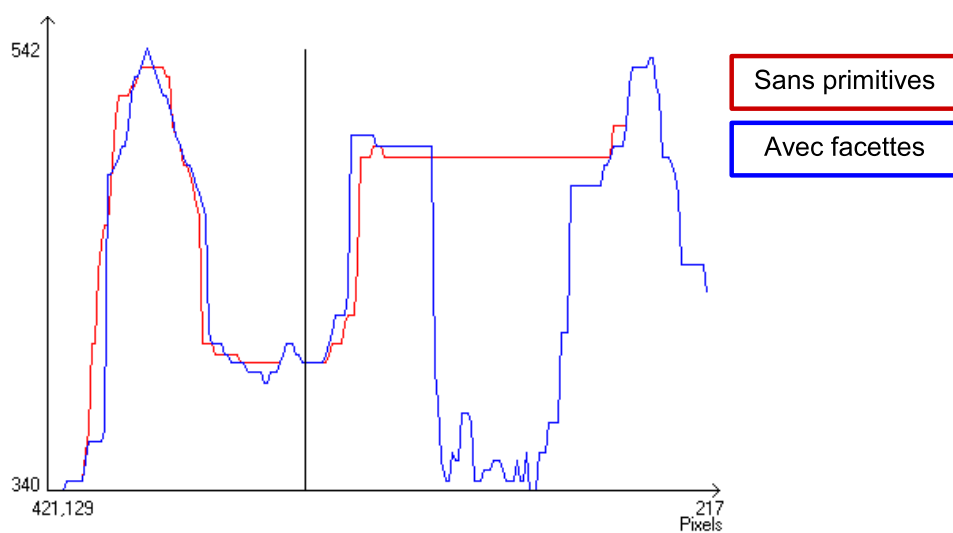
Les facettes extraites



Sans primitives



Avec facettes

FIG. 7.15 – Exemple d'optimisation avec les facettes ($K = 0.5, C_f = 0.5$)

aux bords des bâtiments dus à des façades 3D mal estimées et par lesquelles on force le passage de la coupe. Ce problème pourrait être résolu par la modélisation du sol grâce au réseau routier. La nappe du sol serait alors mise en concurrence avec les facettes extraites (cf. figure 7.16).

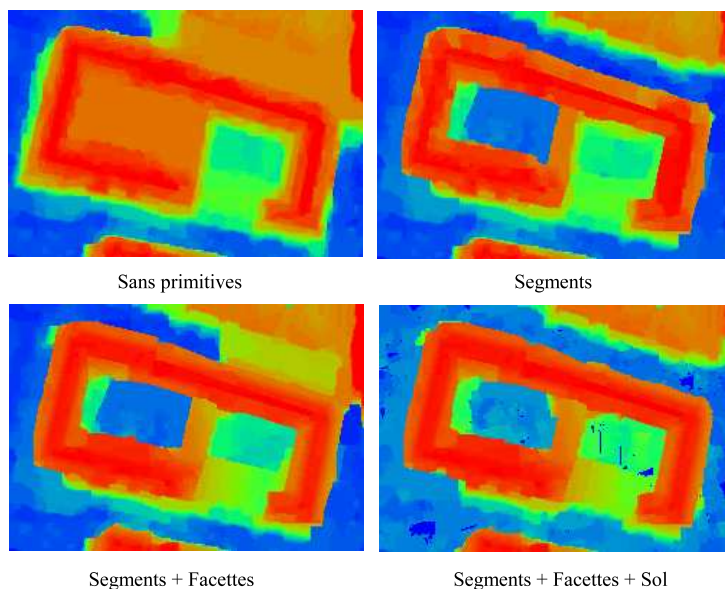


FIG. 7.16 – Exemple d'optimisation avec le sol

Les évaluations de l'optimisation globale contrainte par des primitives 3D seront présentées dans la partie V

7.3.8 Conclusions

On a présenté dans cette étude un algorithme de modélisation de surfaces 3D permettant d'exploiter au mieux la réalité des données dont on dispose. Le problème a été formulé comme un problème de minimisation d'énergie, résolu par une méthode d'optimisation globale basée sur les coupes de graphes. Notre apport consiste à introduire l'information vectorielle fiable extraite et des données externes pour générer un MNE hybride. Ce MNE hybride raster/vecteur combine la grille régulière correspondante au MNE initial et l'information vectorielle correspondante aux primitives 3D fiables extraites.

7.3.8.1 Avantages

L'avantage de notre approche est la gestion explicite des discontinuités par le biais des segments 3D extraits à partir des images. Les facettes 3D sont utilisées pour régulariser les surfaces planes et permettent d'avoir une information supplémentaire sur la pente des toits. La complémentarité de ces deux primitives permet de bien contraindre le problème d'optimisation.

On a proposé également une gestion du graphe 3D par nappes en assurant la prise en compte des magnitudes des discontinuités lors de la coupe et qui permet un gain de temps considérable en diminuant le nombre de nœuds du graphe. Le problème de l'unicité de l'appariement a également été traité. Ce problème est délicat en stéréovision surtout en visée quasi-verticale. On assure la condition d'unicité en présence de primitives 3D.

La formulation proposée permet également de confronter plusieurs hypothèses de facettes pour une même région, les facettes pouvant provenir de plusieurs hypothèses d'appariement ou bien de processus d'extraction complètement différents. Dans ce cas, l'algorithme permet de choisir la meilleure facette en respectant les contraintes appliquées pour la régularité de la surface finale et en évitant le passage intempestif entre les différentes facettes concurrentes.

Un autre avantage de cette méthode est qu'elle est indépendante du processus d'extraction des primitives 3D. La méthode a été présentée dans le cadre de notre étude dans un contexte satellitaire, mais elle reste applicable avec tout type d'images : aériennes, haute résolution, faible ou fort B/H à condition d'utiliser les primitives 3D adéquates dans chaque cas et d'adapter les processus d'extraction. L'algorithme peut aussi être facilement étendu à un contexte multi-vues puisque la coupe minimale est réalisée à l'intérieur du cube de corrélation en 3D, donc l'extension est immédiate.

Enfin, l'algorithme permet à tout moment de remonter à l'information vectorielle correspondante à une facette retenue lors de la coupe afin de représenter la surface finale hybride proposant différents niveaux d'interprétation (raster/vecteur).

7.3.8.2 Inconvénients

Une limitation de cette approche est sa dépendance à la fiabilité des primitives. En effet, la fiabilité des plans de facettes 3D est indispensable pour assurer une bonne estimation des pentes de toit et donc une bonne modélisation de la surface. Par contre, les problèmes de délimitation des facettes ne sont pas critiques et peuvent être résolus soit en utilisant des segments 3D, soit en modélisant le sol. Ce problème de dépendance ne se pose pas pour les segments 3D puisqu'on utilise uniquement leurs positions planimétriques, fiables, pour créer les couloirs de passage.

On relève également les problèmes de discrétisation des facettes qui apparaissent sur la surface finale. Un compromis sera à trouver entre la précision de la reconstruction, en choisissant un pas de discrétisation adéquat, et le temps de calcul.

7.3.9 Perspectives

Pour l'étape d'optimisation globale, on a choisi, par manque de temps, et afin d'adapter à moindre coût cette formulation énergétique à notre contexte, de discrétiser les primitives 3D dans le cube. On envisage, dans la suite de ce travail, de créer un graphe 3D hybride qui serait composé de voxels, mais également de nœuds correspondant à aux primitives vectorielles. *part :ResEval* Dans le cadre de cette thèse, on ne s'est pas intéressés à une modélisation fine du sol. Dans un contexte pratique à l'IGN, on envisage l'introduction des limites cadastrales, qui serait la méthode la plus adaptée et la plus fiable pour le processus d'optimisation globale.

Cinquième partie

Résultats et évaluations

Chapitre 8

Plateforme d'évaluation de MNE

8.1 Introduction

La dernière partie de notre travail de thèse est de définir un protocole d'évaluation pour les données altimétriques obtenues à partir des simulations PLÉIADES. Même si l'étude a une portée plus générale, on ne s'intéressera, dans notre cas, qu'aux MNE calculés automatiquement à partir de couples stéréoscopiques d'images satellitaires sub-métriques. Les enjeux principaux concernant l'acquisition et l'utilisation de ces données altimétriques à haute résolution se situent au niveau des milieux urbains et des objets du sur-sol (bâtiments et végétation). On proposera, dans un premier temps, un état de l'art des indicateurs de qualité utilisés dans la littérature. Dans un deuxième temps, on définira les indicateurs choisis dans notre cadre d'application avec des exemples de résultats et d'évaluations utilisant la plateforme.

8.2 Évaluation de MNE : État de l'art

On pourra séparer les indicateurs de qualité d'un MNE en deux parties : la qualité intrinsèque et la qualité extrinsèque d'un MNE.

8.2.1 Qualité intrinsèque

Un MNE peut être évalué de manière intrinsèque, sans utiliser une référence externe. Deux méthodes d'évaluation intrinsèque de MNE sont présentées dans cette section.

8.2.1.1 Mise en concurrence de plusieurs stratégies

Dans [GC00], l'auteur propose une méthode d'auto-évaluation basée sur la mise en concurrence de plusieurs stratégies de calcul ou, plus exactement, sur la comparaison de deux MNE calculés avec le même algorithme et un paramétrage différent. Sur l'image de différence des deux MNE, les valeurs proches de 0 sont considérées comme des altitudes fiables. Ceci permet de renvoyer une mesure de fiabilité des altitudes du MNE calculé.

8.2.1.2 Comparaison du MNE avec des données image

Dans [Sch00], l'auteur propose une approche basée sur la comparaison entre les lignes de rupture du MNE et les contours de l'image, partant du principe que les discontinuités géométriques (lignes de rupture) correspondent souvent à des discontinuités radiométriques. L'auteur évoque

également la possibilité d'utiliser l'image pour effectuer une validation du MNE sur des critères sémantiques. Les objets détectés dans l'image peuvent être associés à des règles de comportement altimétrique : les surfaces d'eau sont planes et horizontales ou encore les routes sont horizontales perpendiculairement à leur axe et ont une pente inférieure à un seuil le long de leur axe. L'inconvénient de cette approche est que certaines des règles utilisant la sémantique de l'image peuvent être utilisées pour la production du MNE, notamment par les approches basées sur une séparation sol/sur-sol ([Bai97, FC00]). Il devient alors délicat d'utiliser les mêmes critères pour obtenir le résultat et l'évaluer. On écarte donc cette mesure dans notre cas, puisque les contraintes de discontinuités sont utilisées dans le processus d'optimisation globale.

8.2.2 Qualité extrinsèque

La qualité extrinsèque d'un MNE revient à l'évaluer par rapport à une référence. Deux vérités terrain ont été calculées sur la ville d'Amiens ($50cm$, $B/H = 0.2$) et sur Toulouse ($60cm$, $B/H = 0.08$). En pratique, la localisation tridimensionnelle est calculée sans optimisation des paramètres de prise de vue par des points d'appui. Ceci implique que l'erreur altimétrique absolue n'est pas très fiable et que le modèle restitué sera éloigné de la référence absolue. Pour cette raison, des références relatives ont été générées à partir des mêmes couples stéréoscopiques qui ont servi à la reconstruction des primitives 3D. La précision obtenue est alors une précision **relative**. La mesure de la précision géométrique d'un MNE englobe plusieurs sous-problèmes :

- ▶ La mise en correspondance d'un détail du MNE à évaluer avec le détail homologue sur le MNE de référence.
- ▶ Le choix de la mesure de distance entre les détails homologues.
- ▶ Le choix de l'indicateur statistique qui va permettre de qualifier l'ensemble du MNE à partir des distances entre détails homologues.

On distingue trois niveaux de contrôle de la qualité d'un MNE : le contrôle ponctuel, le contrôle linéaire et le contrôle surfacique.

8.2.2.1 Contrôle ponctuel

▶ **Méthode d'appariement et mesure de distance** : Le contrôle ponctuel est la méthode la plus couramment utilisée pour le contrôle de la qualité des MNE. Dans [Abb94], l'auteur en donne une description détaillée. Elle consiste à évaluer la qualité de restitution à partir de points homologues connus à la fois dans la référence et dans le produit restitué. La qualité d'un échantillon restitué est donnée par la distance entre cet échantillon et l'échantillon homologue de la référence. Cette méthode est bien adaptée à l'évaluation de données régulièrement échantillonnées, comme les MNE maillés (raster), où la notion d'homologue est implicite. Elle est cependant bien moins adaptée aux données vectorielles, parce que l'appariement des points caractéristiques n'est pas évident en bords de bâtiments par exemple ou bien parce que les structures homologues issues de la référence et de la restitution n'ont pas la même représentation structurelle en raison de la généralisation.

Comme le montre la figure 8.1, le contrôle ponctuel majore les erreurs en ne mesurant qu'un simple écart altimétrique Δ_z entre deux détails ayant une même position (x, y) . Par défaut, on considère que l'homologue du point M du MNE est le point M'_1 de la référence alors que le point M'_2 est le véritable homologue.

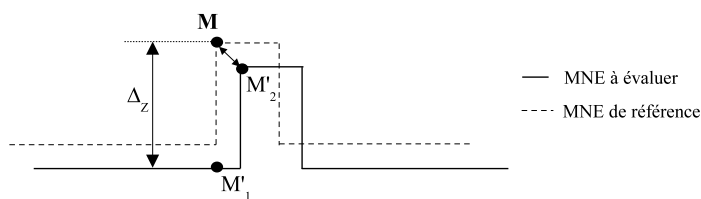


FIG. 8.1 – Comparaison ponctuelle entre un MNE et la référence

► **Choix de l'indicateur statistique :** Pour un ensemble n de points homologues, l'estimation globale de la qualité est souvent donnée par l'erreur moyenne quadratique (EMQ) dont l'expression est la suivante :

$$EMQ = \sqrt{\frac{1}{n} \sum_{i=1}^n \epsilon_i^2} \text{ avec } \epsilon_i = d(P'_i, P_i) \quad (8.1)$$

où d est une distance entre le point P_i de la référence et le point P'_i du produit.

En général, dans le cadre du contrôle ponctuel de MNE, on prendra l'écart altimétrique entre deux pixels ayant la même position planimétrique. Cependant, l'erreur moyenne quadratique n'a vraiment de sens que si l'histogramme des écarts correspond à une loi gaussienne théorique. En supposant que la répartition des écarts réels suit une loi gaussienne, les écarts aberrants sont filtrés en seuillant l'histogramme à une valeur fonction de l'écart-type des mesures. Cette valeur est en général fixée à $\pm 3\sigma$.

Il est également indispensable de calculer le biais sur les mesures en effectuant la moyenne arithmétique des écarts en z entre les mesures et la référence. Ce biais permet de mettre en évidence une erreur systématique correspondant à une translation en z entre les mesures et la référence. Les biais en z apparaissent dans les cas d'évaluations avec une référence absolue. Ce biais est, en partie, dû aux erreurs d'aérottriangulation.

► **Évaluation des erreurs par classes :** La plupart des auteurs ont recours à l'erreur moyenne quadratique comme indicateur de qualité. Cette erreur peut être calculée sur l'ensemble des pixels du MNE. Dans d'autres cas, les écarts ne sont mesurés que sur un sous-ensemble de points caractéristiques [AFW00, GLZ00, KK99]. Dans [SSZ02], les auteurs évaluent et comparent différents algorithmes d'optimisation globale. Ils proposent trois mesures d'erreur qui dépendent de la nature de la région considérée :

- Taux d'erreur dans les régions non occultées : $E_{\bar{O}}$
- Taux d'erreur dans les régions homogènes, non texturées : $E_{\bar{T}}$
- Taux d'erreur dans les régions proches des discontinuités : E_D

8.2.2.2 Contrôle linéaire

Le contrôle linéaire est adapté à l'évaluation des structures linéaires et surfaciques. Il consiste à mesurer l'écart non plus entre points homologues, mais entre des formes ou des primitives homologues A et B. On propose dans ce qui suit, les métriques les plus souvent utilisées pour mesurer l'écart entre deux structures.

► **Distance de Hausdorff** : La distance de Hausdorff entre deux ensembles A et B peut s'écrire :

$$D_H(A, B) = \text{MAX}(\sup \{d(x, B)/x \in A\}, \sup \{d(x, A)/x \in B\}) \quad (8.2)$$

avec $d(x, A) = \inf \{d(x, a)/a \in A\}$

Cette mesure donne la plus grande des plus petites distances entre les deux ensembles A et B. Dans [Abb94], l'auteur utilise la distance de Hausdorff pour le contrôle linéaire planimétrique de base de données vectorielles cartographiques. Toutefois, la mesure de Hausdorff donne une distance maximale entre les deux ensembles. Cette distance est par conséquent très sensible au bruit de localisation des limites des structures. Dans [Hottier 95], l'auteur propose de limiter l'influence de ce bruit, par moyennage de l'ensemble des distances de Hausdorff sur les données délibérément bruitées aléatoirement.

► **Distance de Baddeley** : Afin d'éviter de mesurer une distance maximale entre les deux ensembles, on peut généraliser la distance de Hausdorff [Bad92]. En remarquant que la distance de Hausdorff peut s'écrire sous la forme :

$$D_H(A, B) = \sup \{|d(x, B) - d(x, A)|/x \in A \cup B\} \quad (8.3)$$

On peut généraliser la distance L^∞ de Hausdorff à un espace L^p où $p = [1, \infty[$. Cette nouvelle mesure généralisée peut s'écrire :

$$D_H^p(A, B) = \sqrt[p]{\frac{1}{\text{Card}(A \cup B)} \sum_{x \in (A \cup B)} (d(x, B) - d(x, A))^p} \quad (8.4)$$

Elle traduit alors non plus un écart maximal, mais un écart moyen entre les deux ensembles. Pour $p = 2$, cette mesure s'apparente dans son écriture à l'*EMQ* du contrôle ponctuel. On rencontre dans la littérature d'autres mesures de dissimilarité entre régions [Vin91]. Celles-ci ne sont généralement pas des mesures topologiques et ne sont d'ailleurs pas de véritables distances.

8.2.2.3 Contrôle surfacique ou volumique

De la même manière, on peut généraliser le contrôle linéaire pour évaluer des MNE ou plus généralement des modèles de surfaces 3D. Cette mesure apporterait un plus par rapport au contrôle ponctuel traditionnel notamment pour des reliefs comportant des hautes fréquences. Dans notre cas, on ne s'intéressera pas au contrôle topologique.

Pour le contrôle surfacique, dans [RF99], les auteurs proposent des mesures surfaciques entre des ensembles de régions 2D quelque soit leurs types et leurs structures internes. Ils distinguent deux types de métrique : les métriques géométriques (forme, localisation), qui sont fondées sur l'étude de la différence symétrique entre les deux régions : $a \ominus b = (a - b) \cup (b - a)$ et les métriques structurelles (détection et sur-détection). Les auteurs proposent également des mesures topologies basées sur des graphes d'adjacence entre régions.

À notre connaissance, le problème d'évaluation volumique, on encore 3D n'a pas été traité dans la littérature.

On a présenté dans un premier temps l'état de l'art en évaluation de MNE. Dans la section suivante, la plateforme EvalMNE sera présentée ainsi que tous les niveaux d'évaluations et les métriques mises en œuvre.

8.3 La plateforme

Au cours de cette thèse, on a développé une plateforme d'évaluation de MNE, EvalMne. Ce logiciel permet d'effectuer des évaluations intrinsèques et extrinsèques du MNE. On propose trois niveaux d'évaluations différents : un premier niveau pixel à pixel où les MNE calculés et les MNE de référence sont rasterisés, un deuxième niveau raster/vecteur où le modèle de référence dont on dispose est en vecteur. Le plus haut niveau permet d'effectuer des évaluations vecteur-vecteur entre deux modèles de bâtiments en 3D. Pour les évaluations mettant en jeu des primitives vectorielles 2D ou 3D, la plateforme permet également de mesurer l'exhaustivité de la reconstruction de ces primitives 8.3.1. Cette mesure ainsi que les trois niveaux d'évaluations et leurs indicateurs de qualité respectifs seront détaillés dans ce qui suit.

8.3.1 La mesure d'exhaustivité de la reconstruction

L'évaluation des taux de détection et de sur-détection des primitives par rapport à la référence permet d'estimer l'exhaustivité de la reconstruction des primitives 2D ou 3D. Pour cela, on utilise une technique d'évaluation de primitives proposée dans l'article [MS94]. On notera TP (True Positive), le nombre de facettes de référence qui ont été correctement reconstruites. TN (True Negative), le nombre de facettes de référence qui ne sont pas reconstruites. Enfin, FP (False Positive) sont les facettes reconstruites et qui ne correspondent à aucune facette de référence. On définit alors deux mesures de qualité :

$$\begin{aligned} \text{Taux de détection} &= \frac{TP}{TP + TN} \\ \text{Taux de sur-détection} &= \frac{FP}{TP + FP} \end{aligned} \tag{8.5}$$

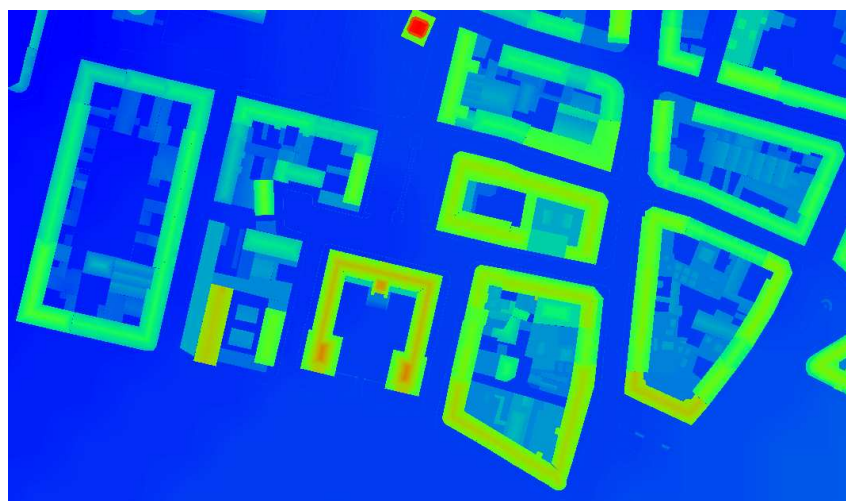
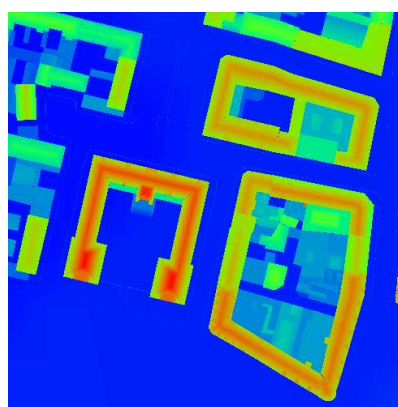
Ces mesures nécessitent une étape de mise en correspondance des facettes calculées et des facettes de référence correspondantes. Cette étape peut être réalisée manuellement par un opérateur, ou bien de manière automatique comme pour l'évaluation des facettes extraites, aussi bien en 2D qu'en 3D (cf. section 6.7.2). Dans notre cas, cette mesure n'a pas été appliquée pour l'évaluation des segments 3D, car les primitives linéiques nécessitent une étape automatique de mise en correspondance plus complexe que celle utilisée pour les facettes.

8.3.2 Évaluation raster-raster

Le premier niveau d'évaluation est l'approche la plus utilisée dans la littérature. Il s'agit de calculer les moyennes et les écarts-types des erreurs entre les mesures et la référence. La figure 8.2 illustre le MNE TRAPU[©] de référence rasterisé à 50 cm.

L'inconvénient de la comparaison pixel à pixel est qu'elle combine les erreurs planimétriques et altimétriques. Un décalage en xy entre la référence et le MNE calculé peut engendrer des grands écarts en z (cf. figure 8.1). Pour décorréler les deux types d'erreurs, les évaluations seront réalisées par classes [SSZ02]. Dans un contexte de scènes urbaines denses, on propose trois classes, choisies de manière thématique, à évaluer séparément : Les bâtiments "B", les bords de bâtiments "BB" qui représentent les régions proches des discontinuités et la classe sol "S".

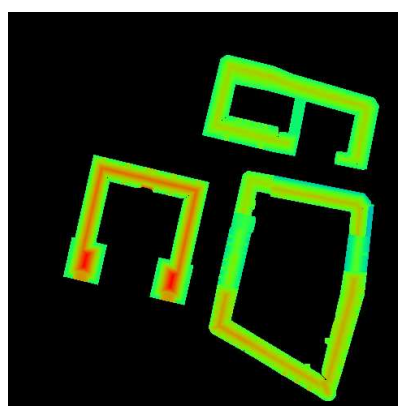
La plateforme fournit un outil interactif pour la saisie des masques de classes. La figure 8.3 montre des exemples de masques de classes pour une zone de test sur le centre-ville d'Amiens.

FIG. 8.2 – MNE TRAPU[®] sur Amiens

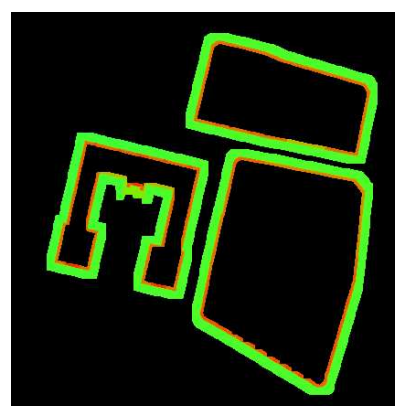
(a) MNE raster de référence



(b) La classe Sol "S"



(c) La classe Bâtiment "B"



(d) La classe Bord de Bâtiments "BB"

FIG. 8.3 – Exemples de masques créés par EvalMne

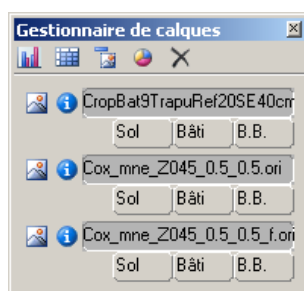
La création des masques peut-être réalisée par différentes méthodes :

- manuellement par un opérateur humain.
- en semi-automatique :
 - par croissance de région à partir d'un germe discriminant choisi par un opérateur (exemple : toit de bâtiment).
 - à partir de l'histogramme du MNE qui fournit des classifications par altitude.
- en automatique :
 - par des opérations morphologiques sur le MNE. Par exemple, le masque de la classe "BB" est obtenu par soustraction du MNE initial à son dilaté.

Le gestionnaire de calques de la plateforme permet d'importer les différents masques et de réaliser les statistiques par classe. Les statistiques sont ensuite calculées sur les différentes classes. On mesure la moyenne, l'écart-type des différences entre le MNE calculé et le MNE de référence et l'EMQ. L'interface des statistiques permet d'importer plusieurs masques et d'exporter les résultats vers un tableur. La figure 8.4 illustre l'interface des statistiques.

En pratique, cette évaluation raster-raster permettra de mettre en évidence l'apport des segments 3D et du sol au niveau des classes "BB" et "S" comme on le verra dans la section 9.2.

A ce niveau d'évaluation, on peut également mesurer localement l'erreur planimétrique. On



(a) Gestionnaire de calques

Classe	Tot Pxl	Min	Max	Moyenne	Ecart-type	Médiane	EMQ	Seuil rejet	% Rejet	Pxl Rest	AR. Moy	AR. ET
Cox_mne_Z045_0.5_0.5.ori[diff]	190071	-129	515	81.8676	76.5206	73	112.061	229	0.234%	189626	80.9494	74.1742
Cox_mne_Z045_0.5_0.5.ori[diff] : Bord de bâtiment	28232	-105	222	110.838	69.7778	144	50.4771	209	0.004%	190063	110.807	69.7641
Cox_mne_Z045_0.5_0.5.ori[diff] : Bâtiment	35314	-105	100	-0.872968	14.8849	1	6.427	44	0.239%	189616	-0.775811	13.3209
Cox_mne_Z045_0.5_0.5.ori[diff] : Sol	72262	-27	198	127.626	48.0319	144	84.0818	144	18.99%	153964	84.2155	26.0765
Cox_mne_Z045_0.5_0.5_f.ori[diff]	190071	-129	515	84.1574	77.2376	76	114.228	231	0.226%	189640	83.258	74.9635
Cox_mne_Z045_0.5_0.5_f.ori[diff] : Bord de bâtiment	28232	-105	222	115.155	69.3632	154	51.8115	208	0.006%	190059	115.112	69.3525
Cox_mne_Z045_0.5_0.5_f.ori[diff] : Bâtiment	35314	-105	100	0.354704	16.0277	2	6.91024	48	0.189%	189711	0.115638	14.8231
Cox_mne_Z045_0.5_0.5_f.ori[diff] : Sol	72262	-26	206	130.132	47.6507	158	85.448	142	21.20%	149768	79.91	20.671

(b) Interface de statistiques ponctuelles

FIG. 8.4 – Interface de statistiques surfaciques (EvalMne)

utilise pour cela le score de corrélation de Pearson qui permet de mesurer la relation linéaire

existant entre deux variables.

$$r_{xy} = \frac{Cov(x, y)}{Var(x) * Var(y)} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$$

avec :

- $Cov(x, y)$ covariance de x et y
- $Var(x)$ variance de x
- $Var(y)$ variance de y

(8.6)

Cette mesure reste toutefois locale et est réalisée le long d'un profil sur les deux MNE à comparer.

Les deux derniers niveaux d'évaluations proposés traitent des primitives de plus haut niveau ; des primitives 2D (les facettes) et des primitives 3D. Les métriques d'évaluation mises en œuvre en 2D et en 3D seront détaillées.

8.3.3 Évaluation raster-vecteur

À ce niveau d'évaluation, notre objectif est d'évaluer un ensemble des facettes 2D (en raster) par rapport à un modèle de référence en 3D. Ceci correspond à un contrôle surfacique du MNE. L'ensemble des facettes peut provenir directement d'un processus d'extraction ou bien, dans notre cas, de la carte de facettes labellisées, retenues lors de la coupe. Le modèle vecteur de référence est projeté dans les images en utilisant un Z-buffer. L'évaluation revient à comparer deux ensembles de comparer deux ensembles de primitives en 2D, en particulier des facettes 2D. La mise en correspondance entre les deux ensembles de facettes 2D est réalisée de manière automatique par un critère de recouvrement planimétrique. On mesure, d'une part l'exhaustivité de la détection (cf. section 8.3.1), et d'autre part on évalue la délimitation des facettes. Les métriques mises en œuvre pour cette dernière partie sont la distance de Baddeley, la longueur de polygonisation et un codage MDL de polygonisation de la facette. Toutefois, ces mesures n'ont pas été utilisées dans les évaluations de nos algorithmes.

Une autre mesure peut également être réalisée sur les ensembles de facettes 2D, qui est la comparaison de leurs pentes respectives. Considérons l'image de différence entre un MNE calculé et un MNE de référence. Pour une région planimétrique donnée correspondant à une facette, la moyenne et la variance des écarts en z permettent d'évaluer la différence des pentes. Par manque de temps, cette méthode n'a pas été testée dans les évaluations.

8.3.4 Évaluation vecteur-vecteur

Ce niveau d'évaluation permet de comparer deux modèles vecteurs en 3D. On ne considère que les primitives 3D représentant des surfaces planes. Cette étape correspond à un contrôle volumique du MNE.

La plateforme permet de réaliser la mise en correspondance automatique entre les facettes calculées et les facettes de référence par des critères de recouvrement planimétrique et de distance volumétrique (cf. section `distVolumique`). Un élagage suivant la distance volumétrique permet d'éliminer les facettes de référence appartenant au sol. Dans un deuxième temps, l'évaluation portant essentiellement sur les toits des bâtiments, la facette supérieure sera retenue.

Comme mesures de d'exhaustivité, on calcule les taux de détection et de sur-détection des facettes 3D. Pour la mesure de précision des facettes 3D, on propose trois métriques : La distance volumétrique, la puissance volumétrique et l'erreur angulaire qui seront détaillées dans ce qui

suit. La figure 8.5 illustre l'interface d'évaluations vectorielles. La plateforme permet de comparer deux modèles simultanément par rapport au même modèle de référence. Les ensembles de facettes appariées entre les différents modèles sont superposés dans une vue 3D afin de pouvoir contrôler visuellement les appariements. On retrouve, dans les statistiques, les métriques mises en œuvre.

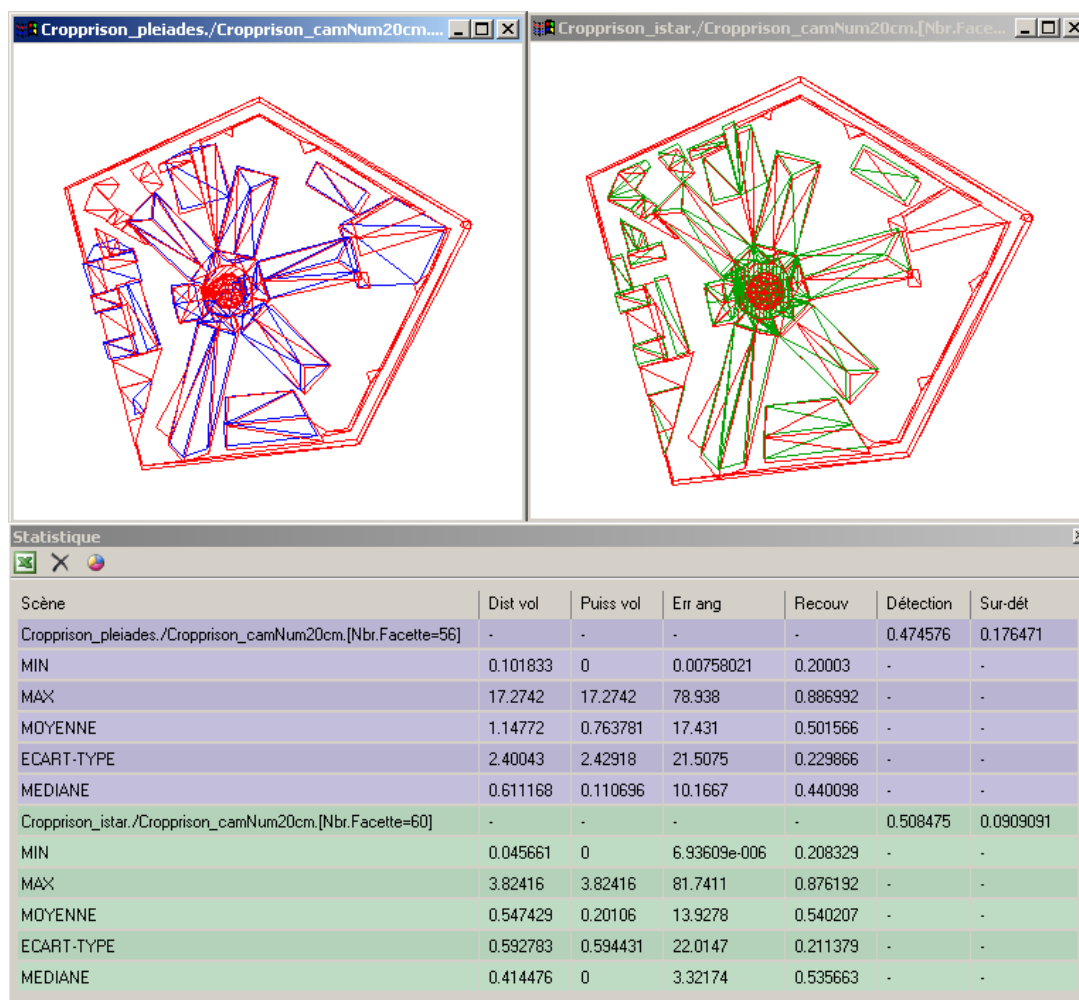


FIG. 8.5 – Interface de statistiques volumiques (EvalMne)

8.3.4.1 Critères d'évaluation des plans 3D

► **Distance volumétrique** : La distance volumétrique permet d'évaluer le plan d'une facette \mathcal{F} par rapport à un plan de référence P_r , en prenant en compte la surface de la facette et la position des deux plans. La figure 8.6 illustre les distances utilisées suivant la position des deux plans.

La distance volumétrique est obtenue comme suit :

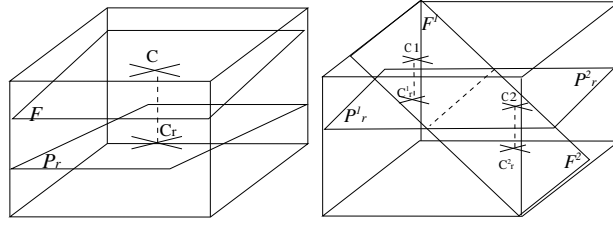


FIG. 8.6 – Distance entre deux plans 3D

$$D(F, P_r) = \begin{cases} |D(C, P_r)| & \text{si } F \text{ n'intersecte pas } P_r \\ \frac{\sum_i A(F^i) \cdot |D(F^i, P_r^i)|}{A(F)} & \text{sinon} \\ A(F) = \sum_i A(F^i) \end{cases} \quad (8.7)$$

avec : $A(F)$ l'aire de la facette F

C : le centre de gravité de F .

D est la distance euclidienne.

L'avantage de cette métrique est qu'elle permet de prendre en compte la position des deux plans. En effet, dans le cas d'intersection des deux facettes, une simple distance entre les centres de gravité des facettes renverrait un score nul. Pour le calcul de la distance volumétrique, la facette F est découpée suivant la ligne d'intersection en plusieurs facettes F^i . La distance finale est égale à la somme des distances des centres de gravité des facettes découpées F^i au plan de référence, pondérées par les surfaces de chaque facette. La distance volumétrique est finalement normalisée par la surface de la facette initiale.

► **Puissance volumétrique :** Cette métrique est calculée de la même façon que la distance volumétrique sauf que les distances sont signées. Elle traduit par conséquent, l'emplacement de la facette évaluée (au-dessus ou au-dessous) par rapport à la facette de référence. Les deux métriques combinées lèvent l'ambiguïté sur la position relative des deux facettes.

► **Erreur angulaire :** Pour comparer les pentes des facettes, on utilise simplement l'erreur angulaire entre les normales des plans des deux facettes

$$e(\perp\pi_f, \perp\pi'_f) < \epsilon \quad (8.8)$$

avec

- $\perp\pi_f$ la normale au plan π de la facette f de référence.
- $e(\perp\pi_f, \perp\pi'_f)$ est l'écart angulaire entre les normales des plans pour comparer leurs pentes respectives.

Le problème de cette évaluation est qu'elle ne tient pas compte de la surface des facettes.

8.3.4.2 Problématique de la comparaison vectorielle

Le problème auquel on a été confronté lors de la comparaison vectorielle est la différence de niveau de généralisation entre modèles. Les modèles 3D proviennent de différentes chaînes de traitement. Suivant le degré de généralisation choisi, les modèles peuvent être très différents, notamment en terme de nombre de primitives utilisées, du niveau de détail recherché, de la présence ou pas de superstructures (chiens-assis, cheminées ...), du découpage des facettes. Cependant, la topologie est généralement respectée entre les différents modèles. D'après ces constatations, il est nécessaire d'avoir le même niveau de généralisation pour comparer deux modèles vectoriels. En pratique, la plateforme donne la possibilité de modifier, de manière interactive, les modèles 3D de référence, par fusion et suppression de facettes, pour éliminer des superstructures qui ne sont pas traitées dans notre cas, ou encore des façades. Des outils de découpage automatique en 3D par opérateur humain ont également été développés pour pouvoir focaliser en 3D sur une zone de test précise.

8.4 Bilan

La plateforme développée propose différents niveaux d'évaluation, en fonction des données dont on dispose. Les travaux dans la littérature portent essentiellement sur le contrôle ponctuel et surfacique. Un certain nombre de métriques ont été mises en œuvre dans la plateforme. Le premier apport se situe au niveau des évaluations raster où l'on dispose d'un outil interactif et intuitif pour la création de masques de classes à partir d'un MNE de référence raster afin de décorrélérer les erreurs planimétriques et altimétriques lors des évaluations.

Le deuxième apport se situe essentiellement au niveau des évaluations 3D où l'on propose deux métriques volumétriques qui prennent en compte la surface des facettes 3D. Des outils de focalisation et de gestion des modèles 3D ont été également développés.

La plateforme a l'avantage de traiter différents types de données : images, images de labels, MNE $2D^{1/2}$, modèles 3D. Elle constitue un système ouvert auquel de nouvelles métriques pourront être rajoutées et son interface est intuitive.

8.5 Perspectives

La plateforme d'évaluation constitue un premier prototype évolutif et ouvert. L'objectif sera d'enrichir la plateforme avec d'autres indicateurs de qualité. Il est évident que la plupart des indicateurs de qualité définis dans ce chapitre, seront applicables dans un contexte d'utilisation plus large, notamment pour des données altimétriques optiques à très grande résolution ou bien obtenues par Laser ou Radar aéroportés.

Afin de mesurer l'erreur planimétrique entre deux MNE, on envisage de développer des méthodes globales de recalage entre MNE raster. Le score de Pearson permet de calculer localement la relation linéaire entre deux variables. Cette métrique peut être utilisée localement sur plusieurs vignettes afin d'estimer une relation affine entre les deux MNE. On pourrait également utiliser des FFT locales suivies d'une estimation globale du mouvement.

Dans un cadre plus général, on envisage de développer un indicateur de niveau de généralisation d'un MNE. Cet indicateur serait intrinsèque au MNE. L'idée serait fondée sur la comparaison des courbes de niveaux à différentes altitudes du MNE. La plateforme pourra également être enrichie par un contrôle topologique et structurel dans le cas de comparaisons vectorielles.

Chapitre 9

Résultats et évaluations

On présente dans ce chapitre les résultats d'optimisation globale obtenus à différentes résolutions ; 50cm, 60cm et 70cm. Les évaluations sont réalisées au niveau raster par rapport au MNE TRAPU[®] et portent sur l'influence des paramètres k et C_f (cf. Eq 7.14).

9.1 Résultats

9.1.1 Amiens 50 cm - $B/H = 0.2$

Les premiers résultats sont montrés sur la ville d'Amiens. Plusieurs zones ont été testées. Cependant, les images de simulations pléiades dont on dispose et le MNE TRAPU[®] de référence ne présentent qu'une petite zone de recouvrement illustrée sur la figure 9.1. Par conséquent, toutes les évaluations extrinsèques à 50 cm et 70 cm seront réalisées sur cette même zone. La figure 9.1 montre les primitives 3D extraites (segments et facettes) qui permettront à contraindre l'optimisation.

Tout d'abord, les résultats d'optimisation sont présentés avec différents paramétrages. Les évaluations sont réalisées, par la suite, en fonction des deux paramètres intervenant dans l'optimisation : k et C_f .

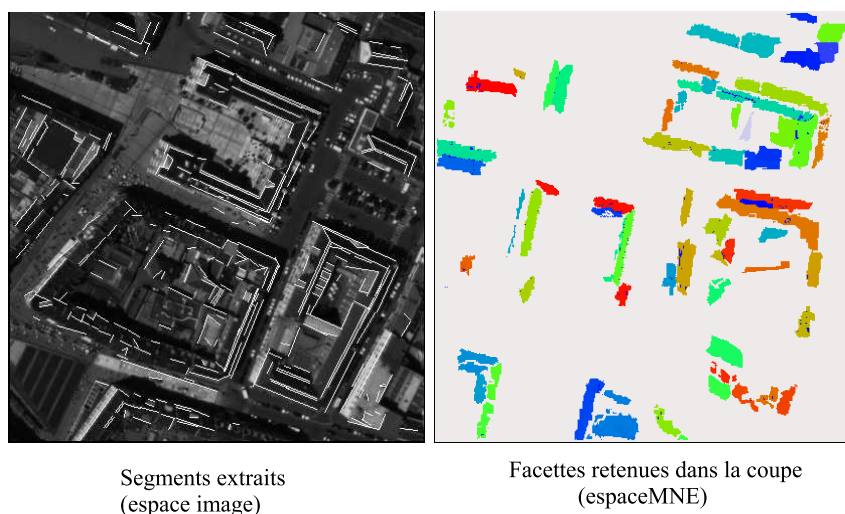


FIG. 9.1 – Amiens 50 cm - Primitives extraites

Sur la figure 9.2, le MNE obtenu sans primitives est complètement lissé à cause de la valeur élevée du paramètre k . L'utilisation des primitives permet de pallier à ce problème et de mieux modéliser les discontinuités et le sol. Cependant, comme le coût fixe C_f est nul, le coût des arêtes est nul au niveau des couloirs de passage et l'altitude maximale est alors retenue pour ces couloirs. Ceci est visible sur les images où la position des segments est repérée par les points d'altitude maximale. En effet, le coût fixe C_f permet de pénaliser d'un même coût minimal toutes les arêtes horizontales. Ce paramètre doit être différent de 0 pour prendre en compte la magnitude des discontinuités.

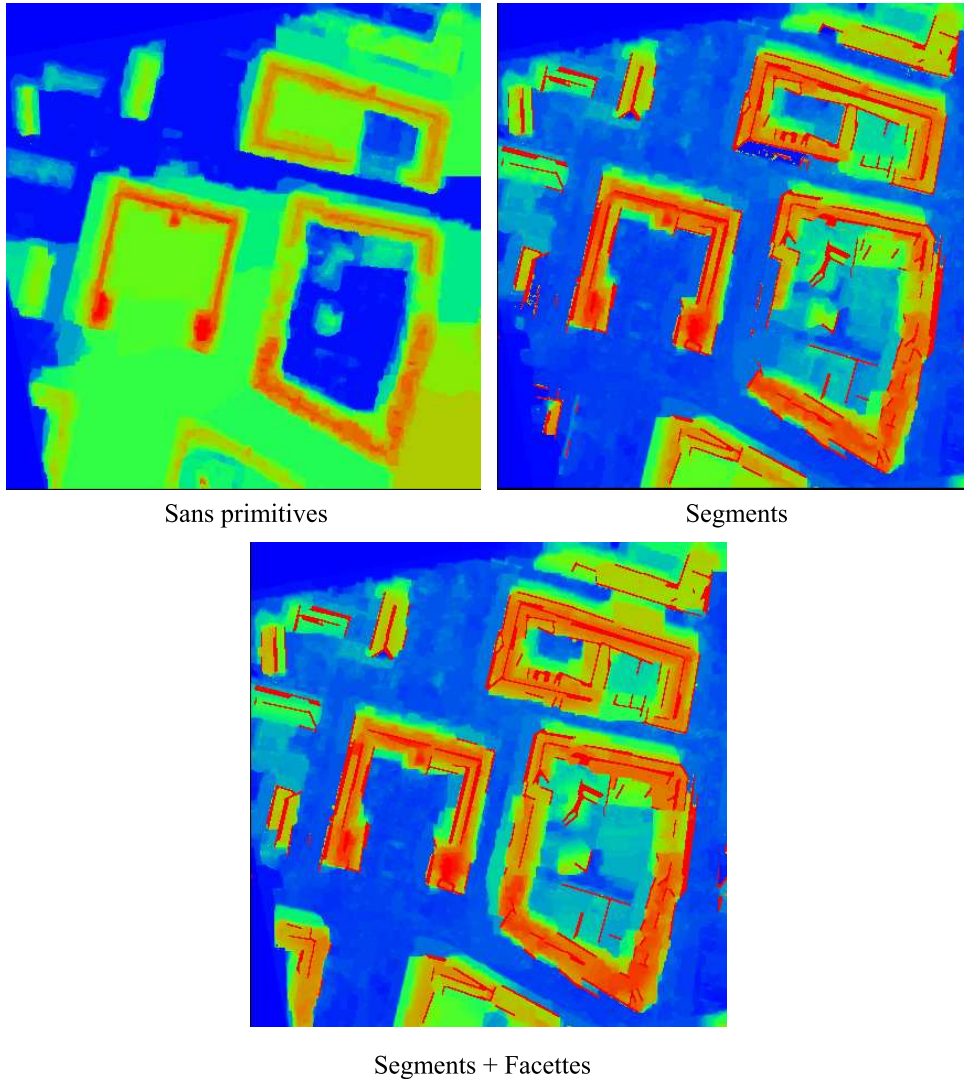


FIG. 9.2 – Amiens 50 cm, $k = 0.5$ – $C_f = 0$

Les figures 9.3 et 9.4 sont obtenues avec des bons paramétrages pour k et C_f . On reviendra sur l'influence de ces deux paramètres dans la section 9.2.2. L'apport des segments est bien visible au niveau de l'optimisation. Les discontinuités sont modélisées et les bords de bâtiments sont régularisés. Ceci permet de mieux modéliser les cours intérieures des bâtiments mais également le sol. Les facettes permettent de régulariser les toits.

L'utilisation conjointe des deux primitives complémentaires permet effectivement de régula-

riser les surfaces planes tout en respectant les discontinuités.

Pour la dernière figure, le sol a été obtenu par classification semi-automatique basée sur une croissance de régions à partir d'un germe, en utilisant le MNE régularisé avec les segments 3D. La modélisation explicite du sol permet de résoudre partiellement le problème de facettes mal estimées ou de facettes n'appartenant pas aux toits.

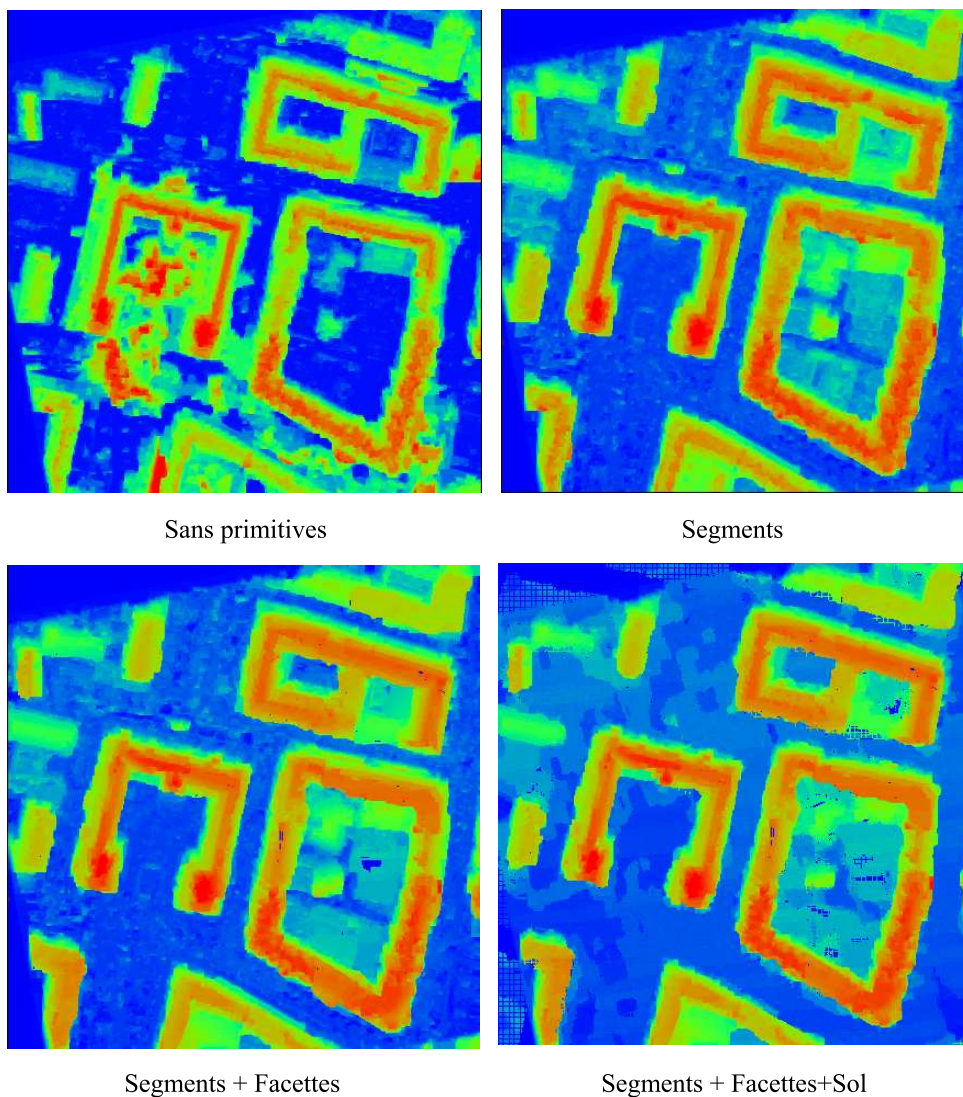


FIG. 9.3 – Amiens 50 cm, $k = 0.2$ – $C_f = 1$

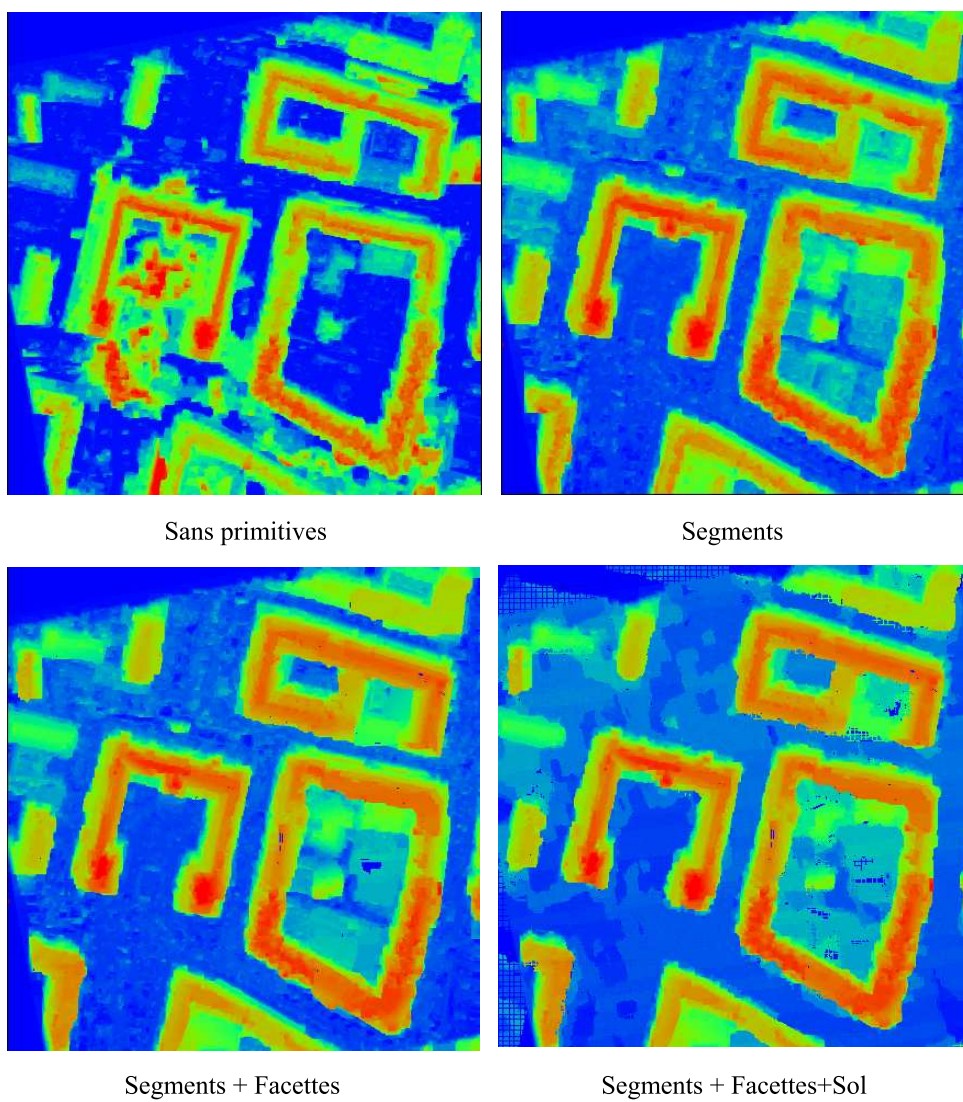


FIG. 9.4 – Amiens 50 cm, $k = 0.1$ – $C_f = 0.5$

9.1.2 Amiens 70 cm - $B/H = 0.2$

Dans cette section, l'optimisation est réalisée sur la même zone du centre-ville d'Amiens avec le même rapport B/H mais à une résolution de 70 cm. Le MNE de référence a été sous-échantillonné à 70 cm.

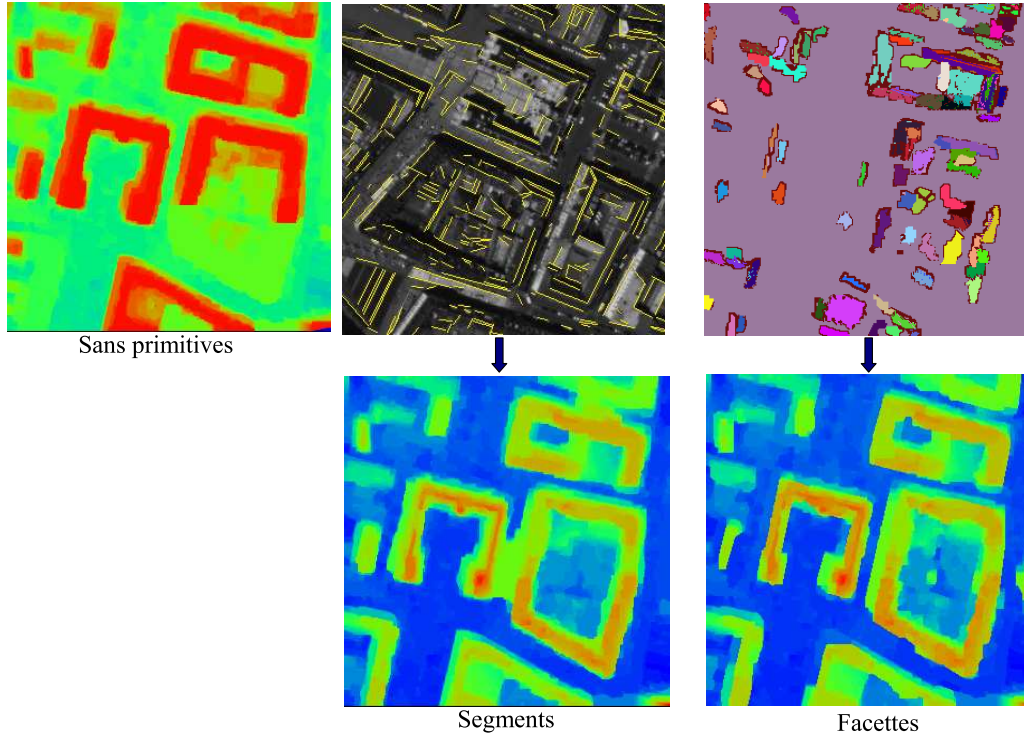


FIG. 9.5 – Amiens 70 cm, $k = 0.5$ – $C_f = 0.5$

Dans le cas de l'optimisation sans primitives, le MNE obtenu est saturé et les toits sont plats. De plus, une partie du bâtiment central est perdue à cause du fort lissage. L'utilisation des segments permet encore une fois de régulariser les bords de bâtiments et de bien modéliser la cour intérieure des bâtiments.

Entre les deux bâtiments centraux, à cause de la présence d'ombres, les segments 3D n'ont pas été détectés sur les bordures de toits. Pour cette raison, les discontinuités relatives à la rue n'ont pas été correctement modélisées. Ce problème est toutefois corrigé par l'utilisation des facettes 3D puisque les couloirs de passage sur les contours des facettes permettent également de modéliser les discontinuités.

9.1.3 Toulouse 60 cm - $B/H = 0.08$

La figures 9.6 et 9.7 montrent les résultats de l'optimisation sur la prison de Toulouse à 60 cm avec un $B/H = 0.08$. À cause du très faible B/H , les facettes 3D extraites ne sont pas fiables en altimétrie. Elles ne seront donc pas utilisées dans le processus d'optimisation globale. Seules les positions planimétriques des segments seront mises à profit.

Les deux figures montrent bien la régularisation des bords de bâtiments grâce aux segments.

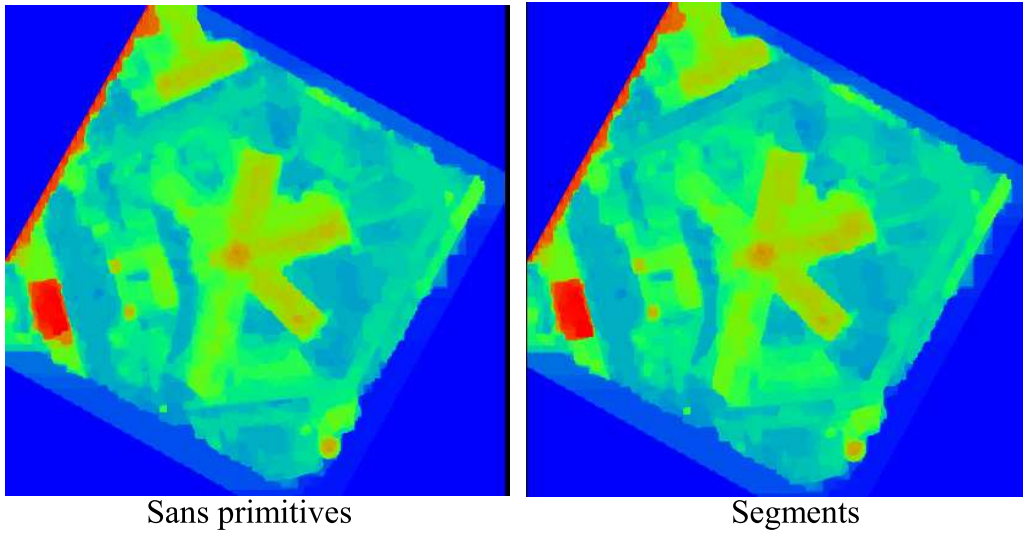


FIG. 9.6 – Toulouse 60 cm, $k = 0.1$ – $C_f = 0.5$

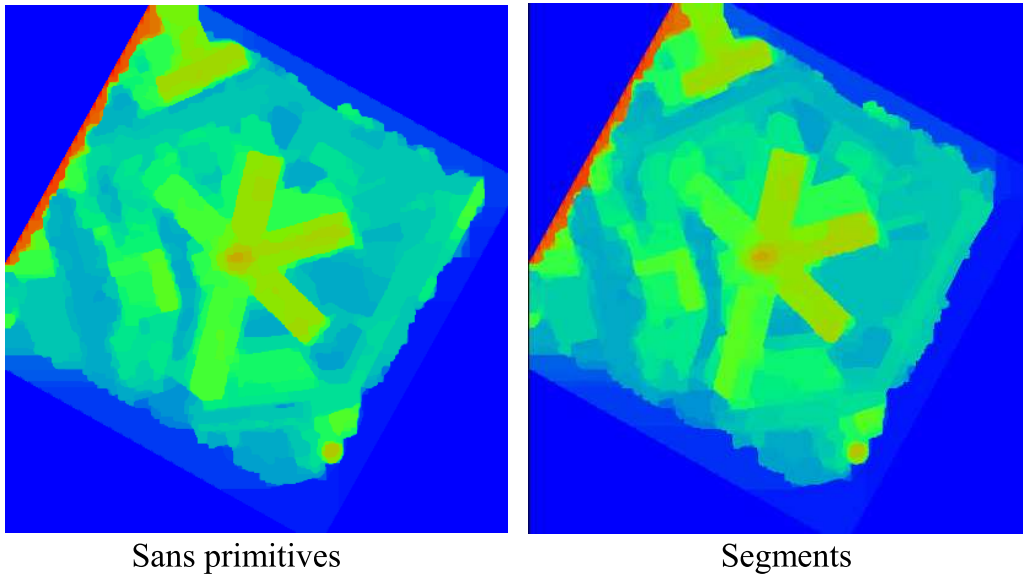


FIG. 9.7 – Toulouse 60 cm, $k = 0.5$ – $C_f = 0.5$

9.2 Évaluations

Cette partie présente les premiers résultats quantitatifs réalisés à partir de la plateforme d'évaluation de MNE. L'évaluation est réalisée au niveau raster (pixel à pixel). Les statistiques portent principalement sur les différences d'altitude entre MNE de référence et MNE calculé. Afin de décorréliser les erreurs planimétriques et altimétriques, les évaluations sont effectuées sur trois classes thématiques : « Bâtiments », « Bords de bâtiments » et « Sol ». Les masques des différentes classes sont générés à partir du MNE de référence.

On mesure par la suite la moyenne, l'écart-type des écarts en altitude et l'erreur quadratique moyenne. L'EQM n'a de sens que si la distribution est gaussienne. La figure 9.8 illustre les histogrammes des différences d'altitude pour les trois classes séparément.

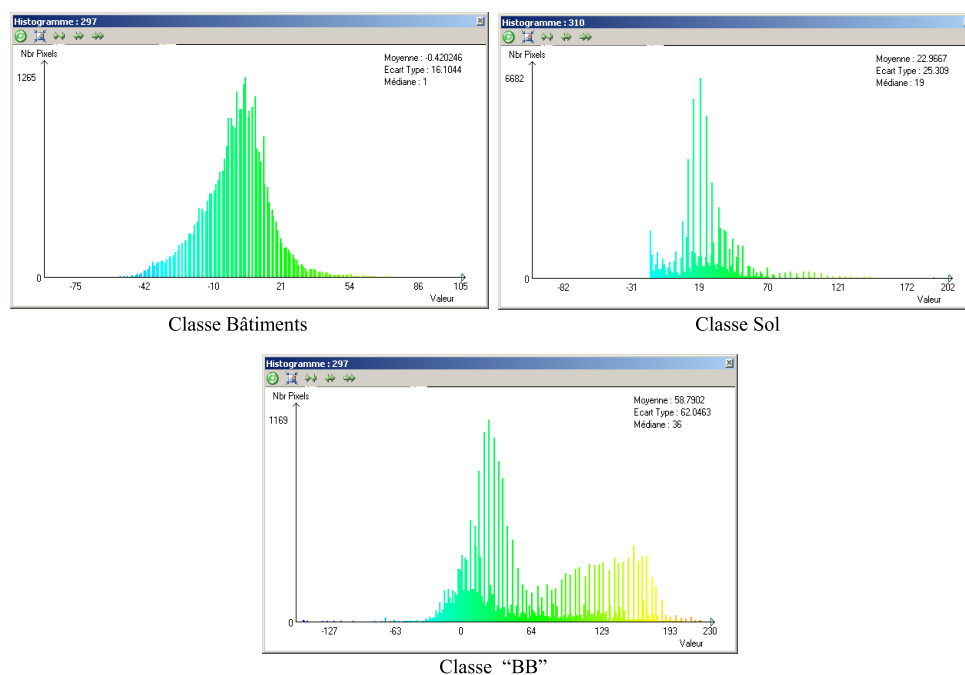


FIG. 9.8 – Histogramme des différences d'altitude par classes

On voit deux exemples de répartitions non gaussiennes qui vont fausser la mesure de l'erreur moyenne quadratique. L'exemple de la classe "S" correspond à la présence d'écarts aberrants en bord d'historgramme, l'exemple de la classe "BB" correspond à une courbe d'erreur bimodale qui met en évidence deux causes d'erreur différentes qu'il conviendrait de séparer préalablement. Dans notre cas, la classe "BB" regroupe effectivement les deux erreurs planimétriques et altimétriques. En pratique, les écarts aberrants sont filtrés en seuillant l'historgramme à une valeur fonction de l'écart-type des mesures, en supposant la loi gaussienne. La valeur du seuil est généralement fixée à 2,6 fois l'écart-type, ce qui correspond à une probabilité de 1% pour une mesure d'être fausse.

9.2.1 Évaluations à 50 cm

La figure 9.9 illustre les différences d'altitude avec le MNE TRAPU®. Les optimisations sont réalisées avec $k = 0.1$ et $C_f = 0.5$. Le coin supérieur gauche des images n'apparaît pas dans le MNE calculé, ce qui se traduit par des gros écarts négatifs par rapport au TRAPU. Inversement, en bas de l'image, les bâtiments n'apparaissent pas dans le MNE TRAPU, ce qui se traduit

par des gros écarts positifs. La figure (a) montre l'impact des segments, la figure (b) montre l'impact conjoint des segments et des facettes. De manière générale, la différence d'altitude varie entre -40 cm et 40 cm sur les toits de bâtiments et une grande partie du sol. Les grosses erreurs apparaissent au niveau des bords de bâtiments où la différence d'altitude est supérieure à 70 cm.

En regardant de plus près les deux images, on remarque que les différences sur les toits sont plus bruitées pour les segments. L'utilisation des facettes permet de régulariser les toits. Par contre, si une facette 3D est mal-estimée, la modélisation des toits est faussée et on obtient des erreurs localisées certes, mais qui sont supérieures à 40 cm sur les toits (cf. figure (b)).



FIG. 9.9 – images de différence d'altitudes $k = 0.1$ - $C_f = 0.5$

Les évaluations qui suivent ont été réalisées sur le bâtiment en forme de "9". La scène a été générée et évaluée avec 10 différents paramétrages. D'après le graphique de la figure 9.10, l'erreur moyenne sur les altitudes de toits est de 20 cm, par contre sur les bords de toit, l'erreur est de l'ordre de 110 cm. Cette dernière classe est difficile à évaluer puisqu'elle regroupe intuitivement les erreurs planimétriques et altimétriques.

L'utilisation des segments améliore les résultats au niveau des classes « BB » et « S » puisque les segments correspondent principalement aux discontinuités des façades. On remarque ensuite, lors de l'utilisation des facettes 3D, que des erreurs sont introduites (des facettes mal reconstruites ou bien des facettes appartenant au sol). Ceci se reflète bien dans les classes "BB" et "S".

Par contre, l'apport des facettes 3D au niveau de la classe « bâtiments » ne se reflète pas dans les calculs de moyenne et d'écart-type puisque la formulation initiale de Roy et Cox est robuste au niveau des toits. La moyenne d'altitude sur la classe "B" reste sensiblement la même. Enfin, les erreurs introduites par les facettes, surtout au niveau des bords de toits sont ensuite corrigées par la modélisation du sol comme facette concurrente aux facettes de toit. Ceci permet de régulariser les bords de bâtiments et d'améliorer l'estimation des altitudes par rapport au modèle de référence.

Les statistiques raster mettent en évidence l'apport des segments 3D pour la régularisation des bords de bâtiments et la modélisation des discontinuités. Cependant, ces statistiques pixel à pixel ne reflètent pas l'apport des facettes 3D essentiellement à cause des problèmes de discrétisation et des décalages éventuels en xy . De plus, la classe "B" est représentée par un nombre faible de pixels 2D et n'est donc pas robuste aux décalages planimétriques.

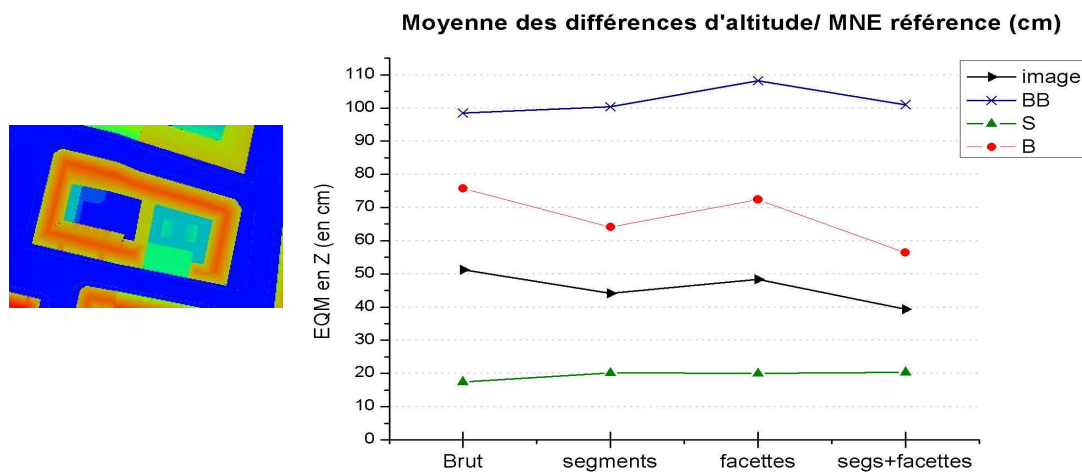
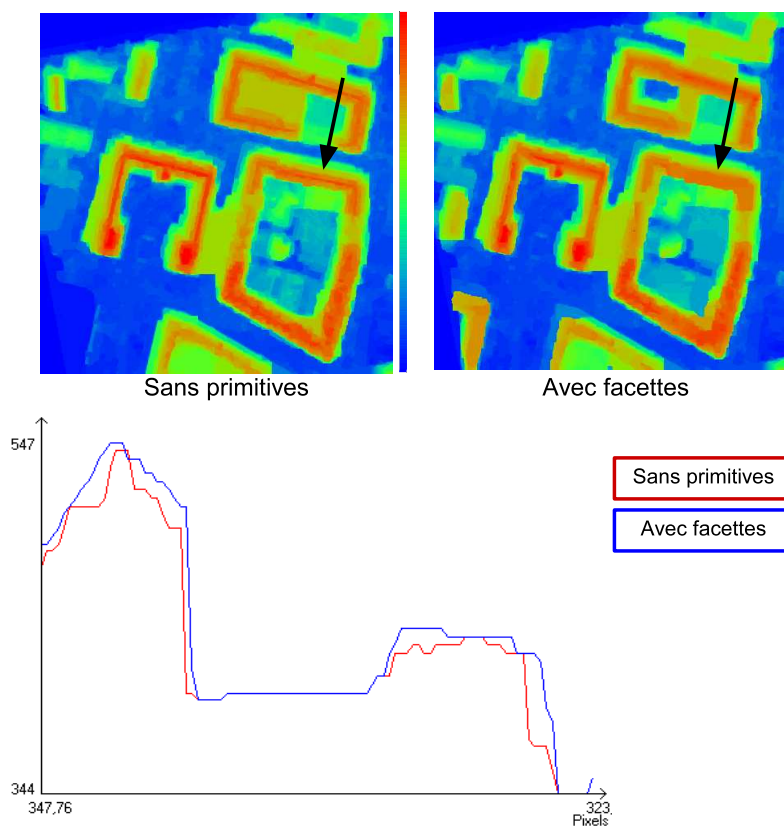


FIG. 9.10 – Moyenne des différences d'altitude par classes

Afin de mettre en évidence la régularisation des toits par des facettes, toujours en raster, on propose de réaliser des profils 1D sur les MNE (cf. figure 9.11).

FIG. 9.11 – Comparaison de profils en présence de facettes ($k = 0.2$ - $C_f = 0.5$)

Les profils en une dimension mettent en évidence la régularisation du toit en présence des facettes mais également les problèmes de discrétisation.

9.2.2 Influence des paramètres : k et C_f

9.2.2.1 Influence du paramètre de lissage k

La figure 9.12 illustre l'influence du paramètre de lissage k . Les évaluations sont réalisées sur la zone de la figure 9.9. Le coût fixe utilisé est $C_f = 0.5$, k variant de 0.1 à 0.5. Les graphiques représentent l'évolution de l'erreur moyenne quadratique en z par rapport au MNE de référence en fonction de k . Les trois classes "B", "BB" et "S" sont évaluées séparément.

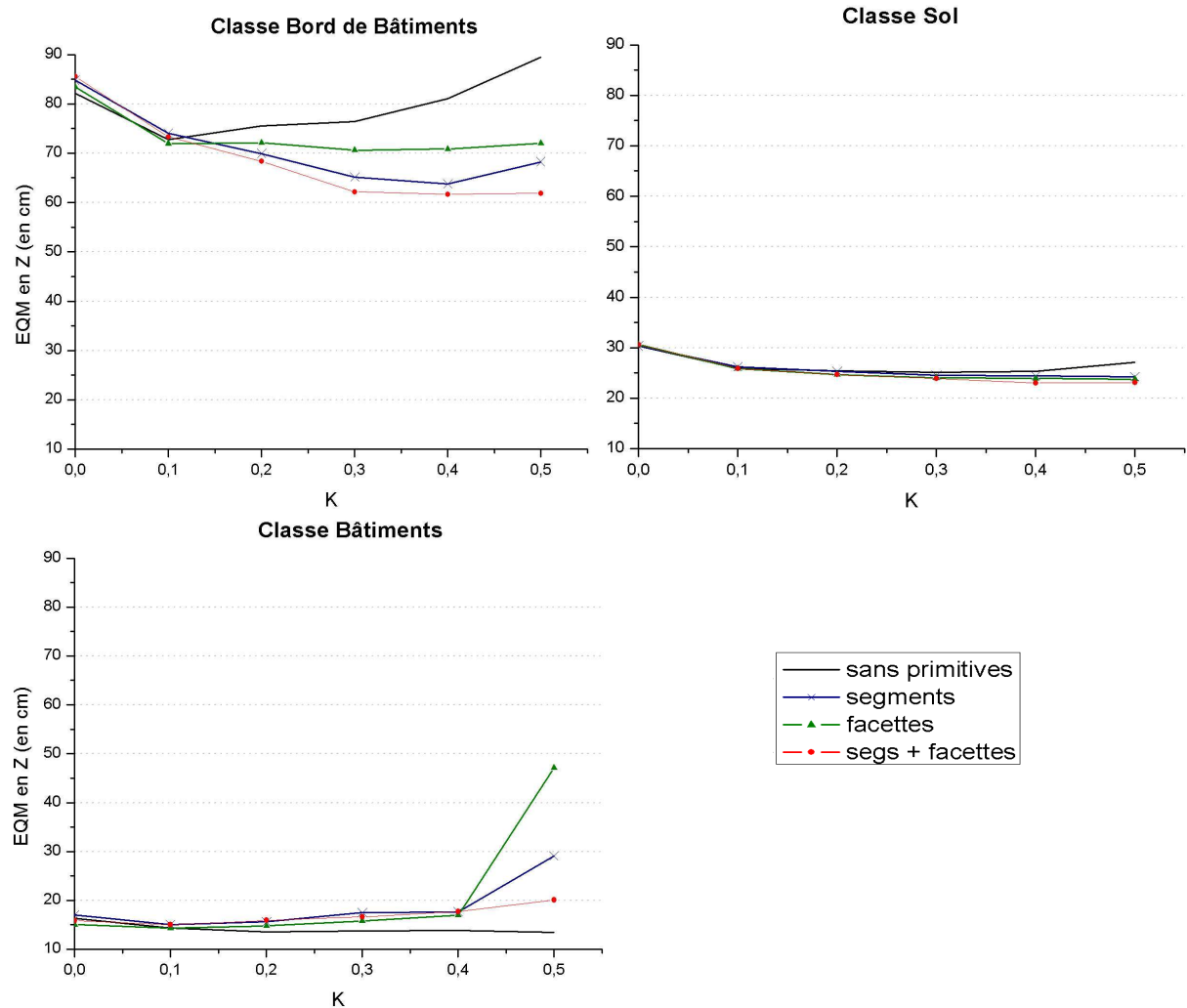


FIG. 9.12 – Influence du coefficient de lissage k - $C_f = 0.5$

Pour la classe "BB", l' EQM varie entre 60 et 90 cm. L'utilisation des primitives améliore l'estimation de l'altitude par rapport à l'optimisation sans primitives. L' EQM diminue en moyenne de 15 cm. On remarque également que les résultats avec les segments sont meilleurs qu'avec les facettes. Ceci était prévisible puisque la délimitation des facettes n'est pas très précise.

L'utilisation conjointe des deux primitives améliore encore plus les résultats. L' EQM de la classe "S" reste stable autour de 25 cm. Aucune modélisation explicite du sol n'a été utilisée.

Enfin, pour les deux classes "BB" et "S", l'utilisation des primitives garantit une estimation robuste de l'altitude pour les bords de bâtiments et le sol pour des forts coefficients de lissage

($k = 0.5$), contrairement à l'optimisation sans primitives. Ce qui traduit une bonne modélisation des discontinuités grâce à une bonne précision planimétrique en contexte satellitaire.

Pour l'estimation des toits de bâtiments, l' EQM en z est stable et se situe autour de 15 cm. Cependant, le résultat sans primitives est meilleur et surtout plus robuste à des fortes valeurs de k . Les résultats sont moins bons en présence de facettes et en cas de fort lissage. On explique ceci, par le manque de précision altimétrique des facettes 3D en contexte satellitaire d'une part, et d'autre part, par le manque de densité des facettes utilisées. L'utilisation conjointe des segments et facettes a tendance toutefois à limiter l'écart en altitude.

À partir de ces trois graphiques, on remarque que l'optimisation est sensible au paramètre de lissage k . Un bon compromis pour la valeur de ce paramètre serait entre 0.1 et 0.3. Elle permet de garder une valeur stable au niveau du sol et des toits de bâtiments tout en améliorant nettement la modélisation des bords de toits.

9.2.2.2 Influence du coût fixe C_f

Les tableaux de la figure 9.13 illustrent l'influence du coût fixe C_f . Les évaluations sont réalisées avec $k = 0.1$ et six paramétrages de C_f variant de 0.5 à 3.

Classe	Sans primitives	Segments	Facettes	Segments + facettes
BB	$72,77 \pm 0,05$	$72,64 \pm 0,82$	$70,73 \pm 0,96$	$71,92 \pm 0,81$
B	$13,68 \pm 0,43$	$15,22 \pm 0,15$	$14,22 \pm 0,15$	$15,15 \pm 0,33$
S	$25,29 \pm 0,36$	$25,53 \pm 0,44$	$25,20 \pm 0,33$	$25,09 \pm 0,54$

FIG. 9.13 – Mesure de l' EQM en cm, Robustesse par rapport au coût fixe C_f

On rappelle la précision théorique pour les MNE (cf. section 2.2.4.1). $\sigma_{xy} = \sigma_z * B/H$. À 50 cm, en supposant une erreur planimétrique d'un 1/2 pixel : $\sigma_{xy} = 25$ cm, $B/H = 0,2$, $\sigma_z = 1.25$ m.

L' EQM pour les toits de bâtiments est de 15 cm. La précision du MNE final est liée au pas de discrétisation du cube. Plus le cube est discrétisé finement, meilleure sera la précision des toits de bâtiments. Dans notre cas, on a choisi un pas de discrétisation en z égal à la résolution de l'image. Ceci permet d'avoir une précision subpixelaire en z au B/H^{eme} près ç.à.d $1/5^{eme}$ de pixel. Par contre, en bords de bâtiments, cette erreur atteint 75 cm. Cette erreur était prévisible puisque cette classe regroupe les erreurs planimétriques et altimétriques et il est très difficile de l'évaluer.

Au niveau du sol, l' EQM atteint 25 cm. Cette valeur reste stable en utilisant les primitives. On rappelle que pour ces évaluations, le sol n'a pas été modélisé de façon explicite. On s'est intéressé dans le cadre de cette thèse à estimer uniquement les facettes de toits.

Le tableau 9.13 illustre également la robustesse de l'optimisation au paramètre C_f puisque l'écart-type est très faible sur les trois classes.

► Évaluation du temps de calcul :

L'optimisation globale a été testée sur une image de taille 191*286, avec un graphe en 6-connexité. La figure 9.14 montre des évaluations du temps de calcul suivant des valeurs croissantes du coût fixe C_f . Elle illustre également, la différence entre l'optimisation globale sans primitives et avec primitives.

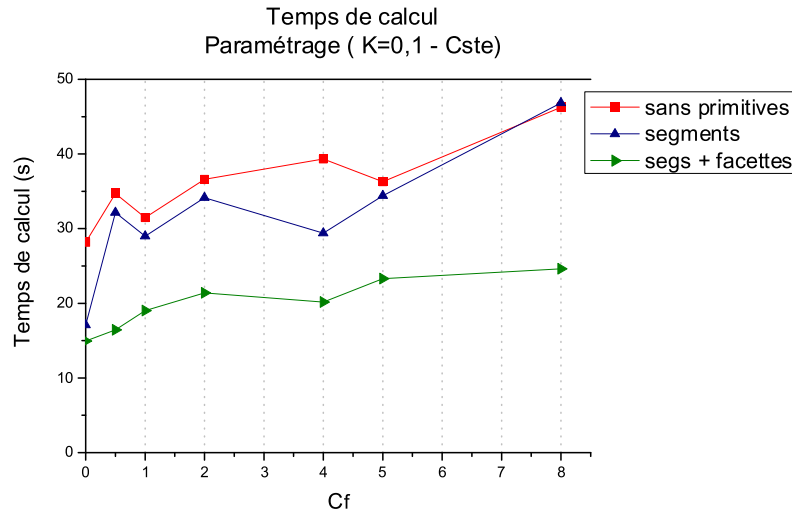


FIG. 9.14 – Influence de la constante C_f sur le temps de calcul

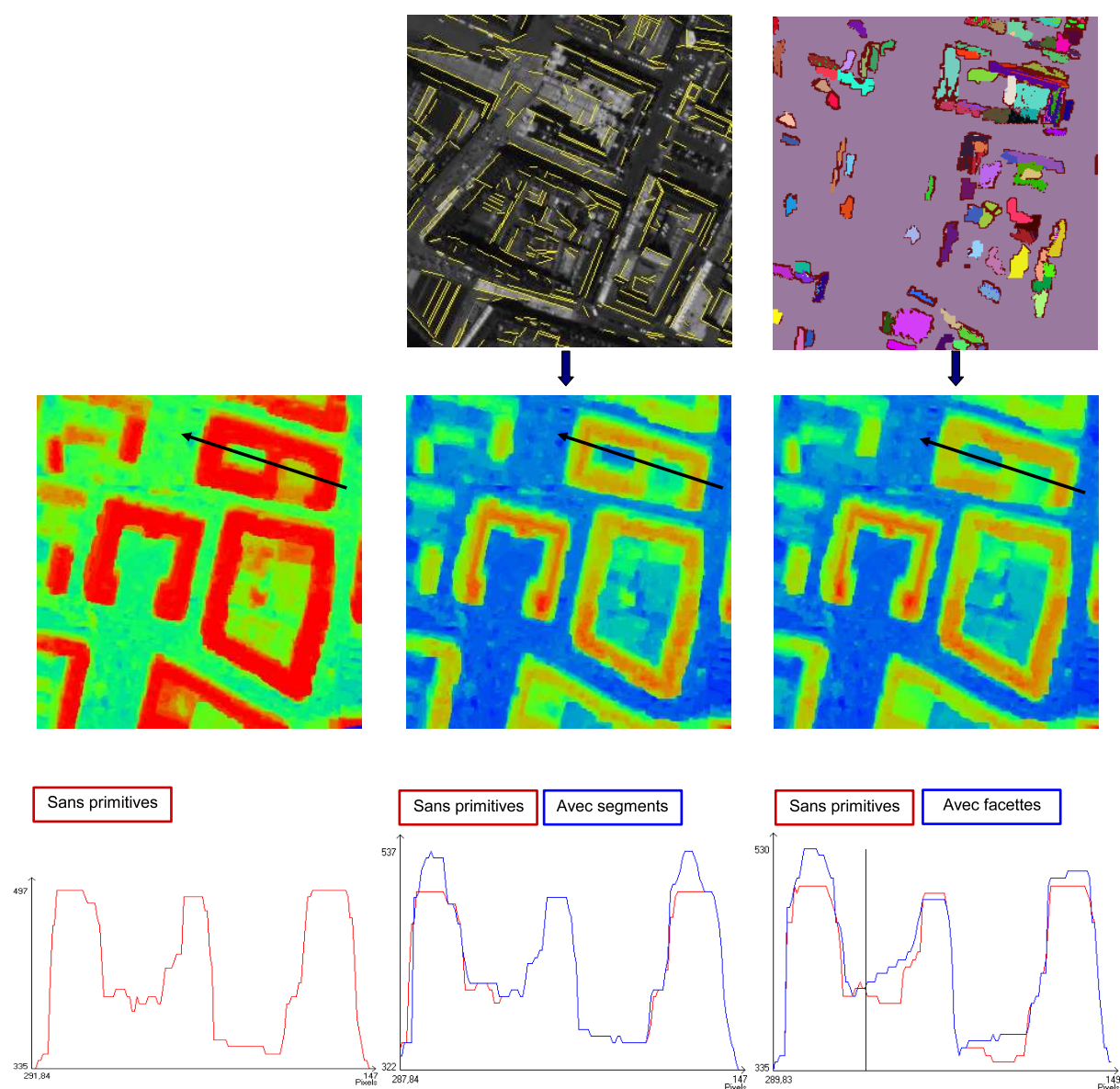
Les temps de calcul sont nettement améliorés en présence de primitives 3D. On obtient un gain de 50% grâce à l'utilisation conjointe des segments et des facettes par rapport à l'optimisation sans primitives.

On voit aussi que pour toutes les optimisations, le temps de calcul croît quand la constante croît. En effet, le coût fixe C_f a pour effet de pénaliser les discontinuités d'un même score minimal. Si C_f augmente, les valeurs des capacités d'occultation C_{occ} sont de l'ordre des capacités de disparité et ne sont plus assez discriminantes. La convergence est alors plus lente. On remarque que pour $C_f = 0$, la convergence est très rapide avec les segments puisque les capacités des couloirs de passage sont nulles (cf. section 7.3.5.0.3).

Dans le cas de l'utilisation des segments et des facettes, le problème d'optimisation étant plus contraint, la convergence est plus rapide et l'algorithme est plus robuste par rapport à la constante C_f .

9.2.3 Évaluation à 70 cm

La figure 9.15 illustre les résultats de l'optimisation à 70 cm avec un $B/H = 0.2$, sans primitives, avec les segments et avec les facettes. Dans le cas de l'optimisation sans primitives, les altitudesaturent et les toits sont plats. L'introduction des segments permet une meilleure modélisation des toits et les bords de bâtiments sont plus réguliers. Le résultat avec les facettes n'est pas satisfaisant. En effet, à 70 cm, les facettes estimées ne sont pas précises en altimétrie ce qui est perceptible à la modélisation de la surface finale. De plus, les profils mettent en évidence les problèmes de discrétisation.

FIG. 9.15 – $k = 0.1$ $C_f = 1.5$

Les résultats quantitatifs à 70 cm ne seront pas donnés à cause d'un fort décalage planimétrique par rapport au MNE de référence qui fausse les statistiques. Une étape préliminaire de recalage 2D des MNE est indispensable.

9.3 Conclusions

D'après les différentes évaluations, on peut conclure que les segments permettent de régulariser les bords de bâtiments et de bien modéliser les discontinuités. L'optimisation n'est pas très sensible à la précision altimétrique de ces primitives puisqu'on n'exploite que les coordonnées planimétriques qui sont fiables. Par contre, la régularisation des toits est très sensible à la qualité des facettes 3D. Dans un contexte satellitaire, on manque malheureusement de précision altimé-

trique pour les surfaces planes, et le fait d'obliger la coupe à passer par ces facettes fausse la modélisation de la surface finale. Mis à part le problème de précision altimétrique des facettes qui apparaît essentiellement à la résolution de 70 cm, on obtient une bonne précision pour le MNE final qui est de l'ordre du B/H^{eme} de pixel. Les évaluations ont permis de mettre en évidence l'apport des primitives dans le processus d'optimisation.

Le deuxième problème rencontré, est le problème de discrétisation des primitives qui est visible sur la surface 3D finale. On a choisi cette méthodologie par manque de temps et pour adapter à moindre coût la formulation de Roy et Cox à notre contexte. Pour s'affranchir de ce problème, on envisage de construire un graphe 3D hybride constitué des voxels du cube et de nœuds correspondants aux différentes primitives.

Chapitre 10

Conclusions

Ce travail de thèse fait partie d'un projet plus global d'exploitation des images satellitaires submétriques pour l'interprétation de scènes urbaines. Cette étude a été initiée afin d'évaluer la modélisation et l'extraction de bâtiments dans un contexte satellitaire très haute résolution. Le travail réalisé au cours de cette thèse, a permis de répondre à un bon nombre d'objectifs fixés au départ. Dans un premier temps, un bilan méthodologique sera présenté, en détaillant les contributions, les inconvénients des méthodologies proposées et les perspectives pour améliorer le système. Dans un deuxième temps, un bilan est réalisé dans le cadre du projet Pléiades-HR en fonction des apports, des limitations du système proposé et des améliorations possibles.

10.1 Analyse méthodologique

10.1.1 Contributions

Tout au long de ce travail de thèse, le faible rapport B/H du couple stéréoscopique a guidé le choix des différentes méthodologies. Toutes les approches proposées tirent profit du faible rapport B/H et traitent les images de manière symétrique. En effet, la contrainte photométrique a été utilisée pour l'extraction des segments 3D et le choix des meilleurs demi-plans qui leur sont attachés. Le processus de délimitation de facettes 3D est également basé sur les scores de corrélation. De plus, le faible rapport B/H facilite l'appariement entre images, c'est la raison pour laquelle on a choisi la mise en correspondance des régions dans l'espace image.

Les deux approches d'extraction de primitives 3D traitent les images de manière symétrique en validant les appariements par un critère d'unicité. Enfin, le processus de modélisation de surfaces 3D est basé sur le cube de corrélation de la scène et traite naturellement les images de façon symétrique puisque le volume est calculé dans l'espace objet.

Nos contributions se situent à trois étapes différentes : la détection de facettes par approche région, le processus de modélisation de surfaces 3D hybrides et la plateforme d'évaluation de MNE. Nos différents apports seront détaillés dans les sections suivantes.

10.1.1.1 Détection de facettes planes par approche régions :

La méthode est basée sur la mise en correspondance de régions. L'appariement étant facilité par un faible rapport B/H . On utilise une description multi-échelles des images qui fournit plusieurs niveaux d'interprétation de la scène. Cette description est adaptée au contexte satellitaire puisqu'elle fournit plusieurs échelles de généralisation des bâtiments. La nouveauté, dans

notre approche, est l'appariement global des deux segmentations multi-échelles. La propagation ascendante des contraintes dans les hiérarchies rend les algorithmes plus robustes au bruit et aux sur-segmentations des régions. La validation des appariements se fait par une stratégie descendante pilotée par la mise en correspondance des régions et un critère de planéité des régions. L'approche fournit des primitives 3D **fiables** avec **un faible taux de surdétection**.

10.1.1.2 Optimisation globale contrainte par les primitives 3D

Le problème de modélisation de surfaces 3D revient à un problème de minimisation d'énergie. Notre apport consiste à introduire de manière explicite des informations de plus haut niveau pour modéliser les discontinuités et régulariser les surfaces planes. La formulation de Roy et Cox [RC98] a été adaptée à notre contexte. Le graphe 3D hybride est construit à partir du cube de corrélation sur la scène 3D et des primitives discrétisées. La coupe de capacité minimale dans le graphe fournit une carte d'altitude **dense** de la scène. Les primitives 3D permettent de gérer explicitement les discontinuités et de régulariser la surface 3D par un ensemble de surfaces planes.

L'avantage de ce processus d'optimisation globale est qu'il est indépendant des processus d'extraction des primitives. Il peut être utilisé dans d'autres contextes (images aériennes, multi-vues) en ayant les primitives adaptées en entrée. De plus, l'optimisation est réalisée à partir du cube de corrélation, dans l'espace objet. Ceci a l'avantage d'une part, de traiter les images de manière symétrique et d'autre part de pouvoir étendre la modélisation de surfaces à un nombre quelconque d'images. Enfin, le fait de construire le graphe 3D à partir du cube de corrélation, fournit une précision subpixellaire au MNE final. En choisissant le pas de discrétisation en z , un compromis sera à trouver entre précision subpixellaire recherchée et temps de calcul.

On a également réduit la complexité de calcul en utilisant une modélisation par nappes à l'intérieur du cube, qui diminue considérablement le nombre de nœuds traités et permet également de prendre en compte les magnitudes des discontinuités lors de la coupe finale (cf. section 7.3.3.3.1).

Le processus est également ouvert puisque toute donnée externe peut être utilisée dans l'optimisation pour contraindre le problème, en passant par une étape de discrétisation des données dans le cube.

On a ainsi proposé une méthode originale de modélisation de surfaces 3D basée sur les coupes de graphe et contrainte par un ensemble de primitives complémentaires extraites à partir de l'image.

10.1.2 Inconvénients

10.1.2.1 Extraction de facettes par approche segment

Cette première approche d'extraction de facettes n'est pas adaptée au contexte satellitaire. L'erreur altimétrique des segments est propagée à chaque étape du processus. L'inconvénient principal de cette approche est l'absence de critère de fiabilité pour les segments 3D, accentuée par la grande imprécision altimétrique de ces primitives linéiques. Par conséquent, il est difficile de qualifier les facettes attachées à ces segments. En outre, on relève un fort taux de surdétection de ces primitives qui est essentiellement dû à la méthodologie utilisée. Les segments 3D ont été utilisés pour leur fiabilité planimétrique pour modéliser les discontinuités. Les facettes qui leur sont attachées n'ont pas été introduites dans le processus d'optimisation globale à cause de leur manque de fiabilité.

Toutefois, ces primitives peuvent être utilisées dans les approches structurales pour la reconstruction de bâtiments [FL99] où l'on cherche à construire un graphe de primitives à partir des données, pour l'apparier au mieux à une bibliothèque de graphes de modèles. Ces approches sont moins dépendantes de la fiabilité des primitives 3D.

10.1.2.2 Extraction de facettes par approche région

Ayant exclu la multiscopie du cadre de cette thèse, l'approche proposée est basée sur la mise en correspondance des régions dans l'espace image. Cette approche est difficilement extensible dans le cas d'un triplet d'images. La combinatoire de la mise en correspondance des segmentations multi-échelles serait trop élevée. Le deuxième inconvénient de cette approche est sa tendance à sur-segmenter les régions retenues dans les segmentations optimales. Ceci est dû au manque de fiabilité des plans estimés sur de petites régions. Toutefois, le seuil de taille minimale permet de pallier à ce problème.

Enfin, le point critique de cette approche est le temps de calcul élevé puisqu'elle nécessite le calcul de deux segmentations multi-échelles des images, leur appariement global et la recherche de facettes par coupes dans les hiérarchies. Cependant, le temps de calcul ne constitue pas un point critique dans le cadre de cette thèse

10.1.2.3 Optimisation globale

L'inconvénient de cette approche est sa dépendance à la fiabilité des primitives extraites. Cet inconvénient n'est pas critique par rapport aux segments 3D dont uniquement les coordonnées planimétriques sont utilisées pour modéliser les discontinuités. Par contre, dans le cas des facettes 3D, les primitives planes doivent être fiables quant à l'équation de leurs plans porteurs. La délimitation des facettes n'est pas critique puisqu'elles sont utilisées dans un processus d'optimisation globale qui gère aussi bien les problèmes de sous-détections des primitives que les problèmes d'hypothèses de facettes concurrentes.

Deux autres inconvénients que l'on pourrait reprocher à la formulation énergétique utilisée par Roy et Cox [RC98] est le fait qu'elle ne garantisse pas la contrainte d'unicité et que les parties cachées soient forcément appariées. Le dernier inconvénient n'en est pas un dans notre cas. En effet, on veut obtenir une description **dense** de la scène et par conséquent, les parties cachées doivent être appariées. Cette contrainte est imposée par le produit final. Concernant la condition d'unicité, elle est garantie seulement dans le cas d'utilisation des facettes 3D. Dans les autres cas, l'altitude la plus élevée est retenue.

10.1.3 Perspectives

Pour l'étape d'optimisation globale, on a choisi, par manque de temps et pour adapter à moindre coût la formulation de Roy et Cox à notre contexte, de discrétiser les primitives 3D dans le cube. On envisage, dans la suite de ce travail, de créer un graphe 3D hybride qui serait composé de voxels du cube de corrélation, mais également de nœuds correspondants à des primitives vectorielles. On s'affranchirait ainsi de l'étape de discrétisation des primitives et des artefacts engendrés sur la surface 3D finale. Une autre amélioration du processus de modélisation de surfaces serait d'utiliser des données externes telles que les limites cadastrales ou le réseau routier pour la modélisation du sol. Ceci permettrait de mieux contraindre l'optimisation globale.

D'autres améliorations sont prévues pour l'étape de détection et de reconstruction de facettes. Grâce à la formulation de Roy et Cox, on peut générer par optimisation sans primitives un MNE subpixelaire plus précis que les MNE de corrélation. Par conséquent, l'étape d'extraction de

facettes planes pourrait être réalisée par segmentation en régions du MNE, en utilisant l'altitude et la planéité des régions comme critères. Les facettes obtenues seraient plus robustes et plus fiables que celles détectées à partir de l'espace image.

Enfin, pour améliorer la précision de reconstruction des facettes 3D, on peut envisager une étape de recalage fin des facettes a posteriori. Le cube de corrélation sera discrétisé plus finement en z . Le recalage des plans des facettes sera basé sur la maximisation des scores de corrélation sur la facette 3D.

10.2 Le projet Pléiades

10.2.1 Contributions

Notre travail a permis de développer une chaîne de traitement complète pour la modélisation de bâtiments 3D. Différents processus d'extraction de primitives linéaires et surfaciques ont été proposés. Le MNE hybride obtenu peut être utilisé dans différentes applications telles que le calcul d'ortho-photos, le survol virtuel de villes, ou encore la texturation de paysages, etc.

10.2.1.1 La plateforme d'évaluation EvalMne

Une attention particulière a été apportée à l'évaluation des différentes primitives extraites ainsi qu'à l'évaluation du MNE hybride final. Tout d'abord, une analyse approfondie de l'extraction des segments 3D a été réalisée. On a montré, par le biais de simulations, les précisions théoriques que l'on pouvait obtenir en fonction de l'orientation du segment, du bruit et du rapport B/H du couple. Sur les images réelles, l'exhaustivité et la précision de la reconstruction ont été mesurées. Il en est de même pour la deuxième approche d'extraction de facettes basée sur la mise en correspondance de régions. Les taux de détection et surdétection des facettes ont été mesurés sur différentes zones et à différentes résolutions.

En ce qui concerne l'évaluation du MNE hybride final, on a développé une plateforme d'évaluation qui permet de réaliser des évaluations à différents niveaux : raster-raster, raster-vecteur ou encore vecteur-vecteur, en fonction de la représentation finale du MNE et du modèle de référence dont on dispose. Cette plateforme regroupe les métriques classiques, de contrôles ponctuel et linéaire, utilisées dans la littérature. Trois métriques sont proposées dans le cas de l'évaluation 3D. La plateforme permet également de quantifier la qualité de reconstruction des facettes 2D et 3D en fournissant les taux de détection et surdétection.

L'avantage de cette plateforme est qu'elle est indépendante du contexte satellitaire et peut être utilisée pour différentes évaluations, dans d'autres contextes, au sein du laboratoire.

10.2.2 Perspectives

On envisage de développer la plateforme d'évaluation et de l'enrichir de nouveaux algorithmes tels que le recalage 2D global entre les MNE pour s'affranchir des erreurs planimétriques et de nouvelles métriques d'évaluation 2D et 3D.

Le point manquant à notre étude est l'évaluation des algorithmes à différents B/H . Cette évaluation n'a pu être menée par manque de données disponibles. On envisage donc de tester les méthodes sur des couples à fort rapport B/H . Il suffirait alors, pour la mise en correspondance des régions, d'adapter la fonction de similitude pour prendre en compte l'homographie entre les deux images. Le reste des contraintes étant appliqué par propagation, aucune autre adaptation ne sera nécessaire. Le processus d'optimisation globale ne sera pas modifié.

10.3 Bilan

Les approches à base de primitives ne permettent pas, dans un contexte satellitaire, d'aboutir à des modèles polyédriques de bâtiments. En effet, les détecteurs de primitives souffrent de problèmes de sous-détections et de surdétections qui rendent le problème de reconstruction de modèles 3D génériques de bâtiments très complexe. De plus, la fiabilité des modèles ainsi reconstruits reste faible de par le manque d'images disponibles pour la vérification et la validation des hypothèses de primitives.

Les approches à base de modèles seront plus appropriées, notamment les approches paramétriques qui permettent de modéliser des structures de toits bien spécifiques en utilisant le MNE comme terme d'attache aux données. Étant donnée la faible résolution des images satellitaires, le manque de généralité des modèles paramétriques ne sera pas un gros inconvénient.

L'utilisation de la tri-stéréo permettrait également de valider les hypothèses de primitives et d'améliorer l'exhaustivité de la reconstruction.

Dans ce qui suit, on essaiera de répondre aux différentes questions posées au début de cette étude.

► Quel B/H faut-il choisir ?

Le choix du B/H va essentiellement dépendre de l'application. Pour des applications cartographiques où l'on cherche à identifier des petits objets, on choisira des faibles ou moyens B/H (0.2 - 0.4) qui permettent, tout en conservant une acuité stéréoscopique suffisante ($B/H > 0.1$) de faciliter les appariements entre images pour la reconnaissance d'objets linéaires tels que les segments ou des surfaces planes. De plus, les B/H faibles maximisent la surface de scène vue en stéréo et permettent par conséquent d'augmenter l'exhaustivité de la reconstruction.

En milieu urbain dense, le choix d'un B/H élevé maximise en théorie la précision altimétrique, mais, rend par contre, le paramétrage de l'appariement plus complexe et contraint la performance de calcul (grandes fenêtres de recherche). De plus, les fortes incidences nuisent à l'exhaustivité de la reconstruction (parties cachées, ombres).

► Quels produits pourra-t-on fournir à partir des images des futurs satellites Pléiades ?

À partir des images satellitaires à très haute résolution, on pourra proposer différents types de produits :

- des modèles numériques d'élévation MNE raster, subpixelaire que l'on peut obtenir avec la formulation de Roy et Cox.
- des MNE hybrides raster/vecteur en 3D, qui combinent les primitives 3D fiables, représentées en vecteur et l'information pixelaire sous forme de grille 3D.
- des modèles 3D polyédriques à partir d'une approche basée sur les modèles.

► Quelles sont les précisions obtenues en fonction des différentes résolutions ?

De part l'approche qu'on a adoptée, qui est basée sur les primitives extraites à partir des images, les évaluations ont essentiellement porté sur les images de résolution 50 cm. À 50 cm, les primitives extraites sont fiables en planimétrie mais manquent de précision altimétrique à cause du faible rapport B/H . L'ensemble des primitives extraites n'est pas dense mais a l'avantage

de présenter un très faible taux de surdétection. Les primitives sont par conséquent fiables. La précision du MNE final est de l'ordre d'un 1/2 pixel de l'image. Cette précision dépend du pas de discrétisation du cube de corrélation. Un compromis est à trouver entre précision de reconstruction souhaitée et le temps de calcul.

À 70 cm, la méthodologie choisie est difficile à évaluer, d'une part parceque les primitives surfaciques ne sont pas fiables à cette résolution, ce qui fausse la modélisation de la surface finale, et d'autre part, le MNE de référence à cette résolution présente une grande erreur planimétrique par rapport au MNE calculé ce qui fausse les statistiques.

Toutefois, l'extracteur de facettes planes donne de bons résultats à 70 cm en terme d'exhaustivité et un faible taux de surdétection. Les évaluations ponctuelles sur les toits de bâtiments montrent une moyenne d'erreur de 70 cm qui correspond également à la résolution des images en entrée.

Annexe A

Propagation ascendante de la contrainte épipolaire

A.1 Introduction

Cet annexe présente l'algorithme de la propagation optimisée de la contrainte épipolaire pour tous les nœuds de référence. Il combine deux propagations : une première propagation dans la hiérarchie de référence pour balayer toutes les régions. Pour une région donnée, n'appartenant pas aux feuilles, les nœuds fils auront déjà été traités, ce qui permet un gain de temps considérable. La deuxième propagation est la propagation de la contrainte épipolaire dans la seconde hiérarchie afin d'obtenir les homologues de chaque nœud de référence par cette contrainte.

A.2 Algorithme

Pour chaque région de référence, on récupère les nœuds fils directs (déjà traités) au lieu de redescendre à ses nœuds de base. Pour chaque nœud fils, on marque ses plus hauts homologues épipolaires (cf. eq A.1) dans la seconde hiérarchie et la propagation ascendante dans H_2 se fera à partir des nœuds marqués.

On notera $HEpiPH(X)$ les plus hauts homologues épipolaires d'un nœud X dans la seconde hiérarchie définis comme suit :

$$HEpiPH(X) = \{x \in HEpi(X) \quad et \quad \nexists \quad y/y \succ x\} \quad (A.1)$$

Rappels :

- Pour un élément x de la hiérarchie, les fils de x sont notés $\mathcal{F}(x)$. L'ensemble des nœuds de base constituant x est noté $\mathcal{B}(x)$.
- Le front de propagation ascendante dans une hiérarchie sera noté *front*.

L'algorithme 2 détaille la propagation optimisée de la contrainte épipolaire appliquée à toutes les régions de la hiérarchie de référence.

Algorithme 2: Propagation optimisée de la contrainte épipolaire pour une hiérarchie

Entrée : Deux hiérarchies H_1 et H_2

Sortie : $HEpi(X) \subset H_2, \forall X \in H_1$

Initialisation : Application contrainte épipolaire pour les nœuds de base de H_1

pour $x \in \mathcal{B}(H_1)$ **faire**

 Trouver les homologues de x par contrainte épipolaire : $HEpi(x)$
 front1 $\leftarrow x$;

Propagation ascendante dans H_1 :

1 **tant que** front1 non vide **faire**

 front1 $\rightarrow x$
 $p(x)$ père du nœud x
 si $p(x)$ a déjà été traité **alors**
 go to 1
 sinon
 pour $x_f \in \mathcal{F}(p(x))$ **faire**
 $HEpi(p(x)) \leftarrow HEpi(x_f)$
 Marquage de $p(x)$
 front1 $\leftarrow p(x)$
 pour chaque $y \in HEpiPH(p(x))$ **faire**
 Marquage de y
 front2 $\leftarrow y$

Propagation ascendante dans H_2 :

2 **tant que** front2 non vide **faire**

 front2 $\rightarrow z$
 $p(z)$ père du nœud z
 si $p(z)$ a déjà été traité **alors**
 go to 2
 sinon
 pour chaque $y \in \mathcal{F}(p(z))$ **faire**
 si y n'est pas marqué **alors**
 go to 2
 Marquage de $p(z)$
 front2 $\leftarrow p(z)$
 $HEpi(p(x)) \leftarrow p(z)$

Annexe B

Les estimateurs robustes

B.1 Introduction

Le problème abordé dans cette annexe est l'estimation robuste de paramètres à partir d'un ensemble d'équations d'observations. Le système d'équation est supposé linéaire et largement surdéfini : si on note m le nombre de paramètres et n le nombre d'équations, on a $n \gg m$. Le système représente un ensemble de données composé d'une part de **données justes** (mesures perturbées par un faible bruit de type gaussien), et d'autre part de **données aberrantes**, c'est-à-dire de valeurs arbitrairement éloignées des mesures réelles et ne correspondant à aucune information. Dans ce qui suit, deux familles d'estimateurs seront détaillées : les M-estimateurs et les estimateurs basés sur un échantillonnage aléatoire dont deux approches seront détaillées : le LMS : (Least Median Squares) et le RANSAC : (RANdom SAmple Consensus).

B.2 Les M-estimateurs

La présentation de la technique des M-estimateurs s'inspire de [Rey83], [Ves00], [RL87] et surtout des présentations de Zhang [Zha95] et [XZ96]. Ces références présentent également d'autres techniques d'estimation robuste, telles que la méthode des moindres carrés médians (Least Median of Square ou LMS en anglais) et les R-estimateurs.

B.2.1 Présentation des M-estimateurs

B.2.1.1 Définition

Le principe de la technique est simple : au lieu de minimiser la somme des carrés des résidus comme dans la méthode de moindres carrés habituelle, on cherche à minimiser une certaine fonction ρ des résidus. En reprenant les notations matricielles habituelles, on cherche le vecteur de paramètres X , solution du système d'équation $AX = B$. Le système s'écrit :

$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,m} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ \vdots \\ b_n \end{bmatrix}, \quad (\text{B.1})$$

où A est une matrice (n, m) , X un vecteur de taille m et B un vecteur de taille n (on a toujours $n \gg m$). On note R le vecteur, de taille n , des résidus :

$$R = AX - B = \begin{bmatrix} r_1 \\ \vdots \\ \vdots \\ r_n \end{bmatrix}, \quad (\text{B.2})$$

et on en déduit la fonction de résidus :

$$res(X) = \sum_{i=0}^n \rho(r_i). \quad (\text{B.3})$$

Le vecteur X est solution du système au sens des M-estimateurs s'il minimise la fonction de résidus res (définie équation (B.3)). Afin que la minimisation ait un sens, il faut que ρ soit une fonction définie positive, avec un minimum unique en zéro. Afin d'être moins sensible aux fortes erreurs, la fonction choisie devra être peu sensible aux très fortes valeurs de résidus.

B.2.1.2 Résolution générale

Au lieu de résoudre le problème de minimisation spécifiquement pour chaque M-estimateur, on montre que le problème peut se ramener à une résolution itérative par moindres carrés, avec prise en compte d'une nouvelle matrice de poids à chaque itération. En effet, on a, pour le vecteur résidu correspondant au vecteur X , solution du système :

$$\min \sum_{i=0}^n \rho(r_i) \quad (\text{B.4})$$

où r_i est une fonction des x_j . On en tire, par dérivation par rapport à chaque variable x_j :

$$\sum_{i=0}^n \psi(r_i) \frac{\partial r_i}{\partial x_j} = 0, \quad \forall j \in [1, m] \quad (\text{B.5})$$

où on a noté ψ la fonction dérivée :

$$\psi(x) = \frac{d\rho(x)}{dx}. \quad (\text{B.6})$$

On définit alors la fonction de poids, notée w :

$$w(x) = \frac{\psi(x)}{x} = \frac{1}{x} \frac{d\rho(x)}{dx}. \quad (\text{B.7})$$

On peut alors réécrire l'équation (B.5) sous la forme :

$$\sum_{i=0}^n w(r_i) r_i \frac{\partial r_i}{\partial x_j} = 0, \quad \forall j \in [1, m]. \quad (\text{B.8})$$

Or cette équation est exactement celle obtenue par dérivation du système de moindres carrés :

$$\sum_{i=0}^n w(r_i) r_i^2, \quad (\text{B.9})$$

qui est donc minimum.


Pour trouver la solution au sens d'un certain M-estimateur du système $AX = B$, il suffit donc de résoudre ce système au sens des moindres carrés, après l'avoir pondéré par une matrice de poids. Cela revient donc à résoudre le système $WAX = WB$ avec :

$$W = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & w_n \end{bmatrix}, \quad (\text{B.10})$$

où on a noté $w_i = w(r_i)$. La solution aux moindres carrés de ce système est évidemment :

$$X = (A^t W^t W A)^{-1} W B \quad (\text{B.11})$$

Cependant, cette matrice de poids dépend elle-même des résidus, puisque $w_i = w(r_i)$. La résolution se fait donc par une méthode itérative : à chaque itération (k), on résout le système $W^{(k-1)}AX^{(k)} = W^{(k-1)}B$ par moindres carrés. $W^{(k-1)}$ représente la matrice de poids calculée à partir des résidus de l'itération ($k-1$). Cela peut donc s'écrire sous la forme :

Initialisation de $X^{(0)}$, $k = 1$	
Calcul du vecteur résidus $R^{(k-1)} = AX^{(k-1)} - B$	
Calcul de la matrice de poids $W^{(k-1)}(R^{(k-1)})$	
Résolution de $W^{(k-1)}AX^{(k)} = W^{(k-1)}B$	
si $\ X^{(k)} - X^{(k-1)}\ > \text{Seuil}_{convergence}$	

$k = k + 1$

où $\text{Seuil}_{convergence}$ est une valeur permettant de déterminer la fin de la convergence.

Cette méthode permet de résoudre le système pour n'importe quel M-estimateur de manière relativement performante. La convergence de cette méthode dépend du comportement de la fonction ρ et doit donc être étudiée pour chaque M-estimateur.

B.2.1.3 Propriétés générales

La fonction ψ (définie équation (B.6)) est appelée fonction d'influence. En effet, la fonction ψ mesure l'influence d'une donnée sur les valeurs estimées des paramètres. Dans le cas classique des moindres carrés ($\rho(x) = x^2/2$), la fonction d'influence est $\psi(x) = x$, c'est à dire que l'influence d'une donnée augmente linéairement avec son erreur, ce qui démontre la non-robustesse de cet estimateur. Pour qu'un estimateur soit robuste, il ne faut pas qu'une seule donnée erronée suffise à créer un décalage significatif sur le résultat.

De ces observations, on déduit un certain nombre de propriétés souhaitées pour le M-estimateur "idéal" :

- une fonction d'influence ψ bornée, afin de limiter l'influence des fautes ;
- un minimum unique : cela implique que la fonction de résidus $res(X)$ ait un minimum unique, ce qui revient à imposer que la fonction ρ soit convexe par rapport à la variable X , ce qui est équivalent à $\frac{\partial^2 \rho}{\partial X^2}$ défini positif ;
- lorsque la dérivée seconde $\frac{\partial^2 \rho}{\partial X^2}$ est singulière, le gradient $\frac{\partial \rho}{\partial X}$ est non nul : c'est une nécessité pratique pour permettre d'assurer la convergence du processus de minimisation. Cela évite une résolution par parcours de l'espace des paramètres.

B.2.1.4 Application

Afin d'illustrer les avantages et inconvénients des différents M-estimateurs, trois systèmes d'équations seront traités de manière explicite pour chacun d'entre eux. Le premier système est très simple :

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} [x] = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 20 \end{bmatrix}. \quad (\text{B.12})$$

Ce système, malgré sa simplicité, est représentatif du type de problèmes rencontrés. C'est un système à une inconnue, trois données bruitées et une faute. Il a pour solution $x = 0$ et pour écart-type sur les données juste environ 1. Les deux autres systèmes sont :

$$\begin{bmatrix} 10 \\ 10 \\ 10 \\ 10 \end{bmatrix} [x] = \begin{bmatrix} -10 \\ 0 \\ 10 \\ 200 \end{bmatrix} \quad (\text{B.13})$$

et

$$\begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} [x] = \begin{bmatrix} -0.1 \\ 0 \\ 0.1 \\ 2 \end{bmatrix} \quad (\text{B.14})$$

Ils sont équivalents au premier, à une renormalisation près. Ils seront utilisés pour tester la résistance des estimateurs aux problèmes de mauvaise estimation de l'écart-type a priori sur les données justes. Le premier d'entre eux (B.13) correspond à un écart-type a priori sur les données justes d'environ 0.1 et le second (B.14) à un écart-type d'environ 10. Pour les différents estimateurs, on a tracé les fonctions résidus pour les différents systèmes ($res_{\sigma=1}(x)$ pour le système (B.12), $res_{\sigma=0.1}(x)$ pour (B.13) et $res_{\sigma=10}(x)$ pour (B.14)). Ces fonctions s'écrivent :

$$res_{\sigma=1}(x) = \rho(x+1) + \rho(x) + \rho(x-1) + \rho(x-20) \quad (\text{B.15})$$

$$res_{\sigma=0.1}(x) = \rho(10(x+1)) + \rho(10x) + \rho(10(x-1)) + \rho(10(x-20)) \quad (\text{B.16})$$

$$res_{\sigma=10}(x) = \rho(0.1(x+1)) + \rho(0.1x) + \rho(0.1(x-1)) + \rho(0.1(x-20)) \quad (\text{B.17})$$

Elles ont été tracées sur deux intervalles :

- $x \in [-2, 2]$: pour avoir une vue locale autour de la solution,
- $x \in [-2.5, 22.5]$: pour avoir une vue globale du comportement de l'estimateur sur l'ensemble des données.

Les paragraphes suivants présentent des exemples de M-estimateurs parmi les plus courants. Ils ont été séparés en deux groupes : les L-estimateurs (paragraphe B.2.2) et les M-estimateurs généraux (paragraphe B.2.3).

B.2.2 L-estimateurs

B.2.2.1 Généralités

Les L-estimateurs (notés L_p) constituent un sous-ensemble des M-estimateurs. Leur fonction ρ est de la forme :

$$\rho(x) = \frac{|x|^p}{p} \quad (\text{B.18})$$

Cette famille intègre les deux estimateurs les plus courants : la minimisation par moindre carrés (L_2) et la minimisation par moindre valeur absolue (L_1). Une des caractéristiques de cette famille est d'être indépendante de la renormalisation des équations. En effet, si on multiplie chaque ligne des matrices A et B par λ , le résidu de chaque ligne est également multiplié par λ . Or :

$$\rho(\lambda x) = |\lambda|^p \rho(x), \quad (\text{B.19})$$

d'où

$$\text{res}(\lambda X) = |\lambda|^p \text{res}(X), \quad (\text{B.20})$$

Les fonctions à minimiser ($\text{res}(X)$ et $\text{res}(\lambda X)$) sont proportionnelles. Elles atteignent donc leur minimum au même point : les trois fonctions de résidus ($\text{res}_{\sigma=1}$, $\text{res}_{\sigma=0.1}$ et $\text{res}_{\sigma=10}$) sont équivalentes. Par conséquent, seule la fonction $\text{res}_{\sigma=1}$ sera illustrée dans ce qui suit.

B.2.2.2 Estimateur L_2

B.2.2.2.1 Présentation Cet estimateur, défini dans le tableau B.1 et la figure B.1, est le meilleur estimateur à partir de données bruitées par un bruit gaussien. Son défaut est sa très forte sensibilité aux fautes, puisque la fonction d'influence n'est pas bornée. L'estimateur L_2 n'est donc pas un estimateur robuste.

$\rho(x)$	$\psi(x)$	$w(x)$
$\frac{x^2}{2}$	x	1

TAB. B.1 – Définition de l'estimateur L_2

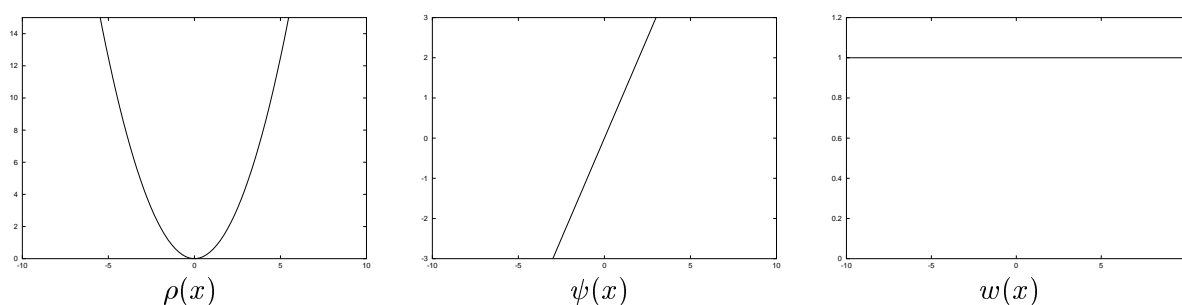
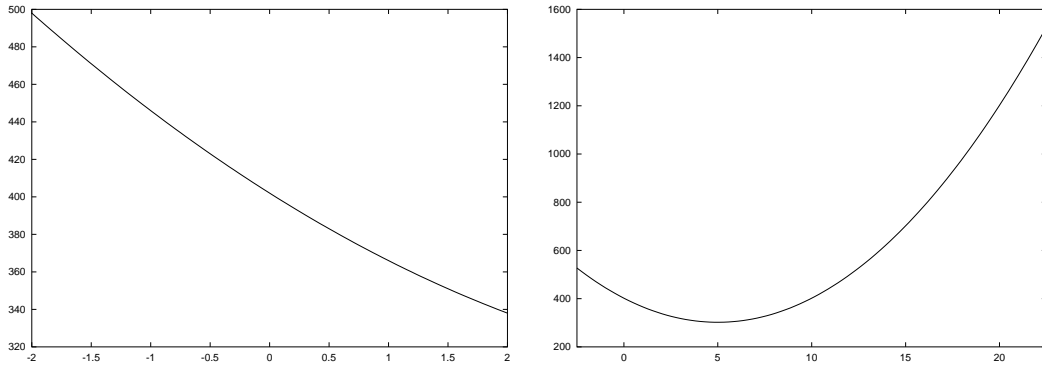


FIG. B.1 – Estimateur L_2 : visualisation

B.2.2.2.2 Application Dans le cas considéré, la solution pour cet estimateur est la somme des données divisée par la somme des poids, ce qui donne $x = 5$ dans tous les cas. On constate bien l'indépendance par rapport à l'écart-type a priori, mais également l'absence de robustesse de l'estimation.

FIG. B.2 – Estimateur L_2 : $res_{\sigma=1}(x)$ sur les intervalles $[-2, 2]$ et $[-2.5, 22.5]$

B.2.2.3 Estimateur L_1

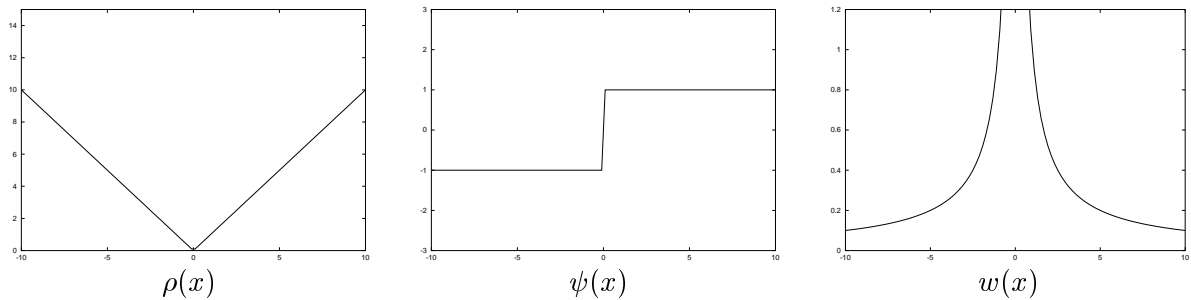
B.2.2.3.1 Présentation L'estimateur L_1 , défini dans le tableau B.2 et la figure B.3, également appelé "estimateur par les moindres valeurs absolues" est très utilisé. $sgn(x)$ désigne la fonction signe (-1 si $x < 0$, 1 si $x > 0$). Il est beaucoup moins sensible aux fautes, mais n'as-

$\rho(x)$	ψ	w
$ x $	$sgn(x)$	$1/ x $

TAB. B.2 – Définition de l'estimateur L_1

sure pas l'existence d'un unique minimum. En effet, la fonction $\rho(x) = |x|$ n'est pas strictement convexe en x , puisque sa dérivée seconde en 0 est non bornée. De plus, le calcul des poids pour la résolution itérative peut poser problème car les poids deviennent infinis quand les résidus sont nuls. Enfin, même si cette fonction réduit l'influence des grandes erreurs, ces données sont tout de même prises en compte puisque la fonction d'influence ψ n'a pas un support borné. Cette minimisation donne cependant un résultat correct tant que le poids des données fausses est inférieur au poids des données justes.

L'estimateur L_1 est donc un estimateur robuste, ne nécessitant pas de connaissances a priori sur l'écart-type des données justes, qui permet de converger en étant relativement peu sensible aux fautes. Cependant, sa convergence et la valeur exacte de la solution finale ne sont pas toujours satisfaisantes. Il existe un algorithme spécifique pour résoudre le système dans le cas de l'estimateur L_1 : l'algorithme de Barrodale [BR72].

FIG. B.3 – Estimateur L_1 : visualisation

B.2.2.3.2 Application Dans le cas traité, la solution pour cet estimateur est la médiane des données. On vérifie bien que l'estimateur est indépendant de la valeur de l'écart-type a priori. La non convexité pose problème : la fonction à minimiser est constante et minimale sur l'intervalle $[0, 1]$ dans tous les cas. L'algorithme peut proposer toute valeur dans cet intervalle. Cependant, quelle que soit la valeur choisie, l'estimation est relativement correcte.

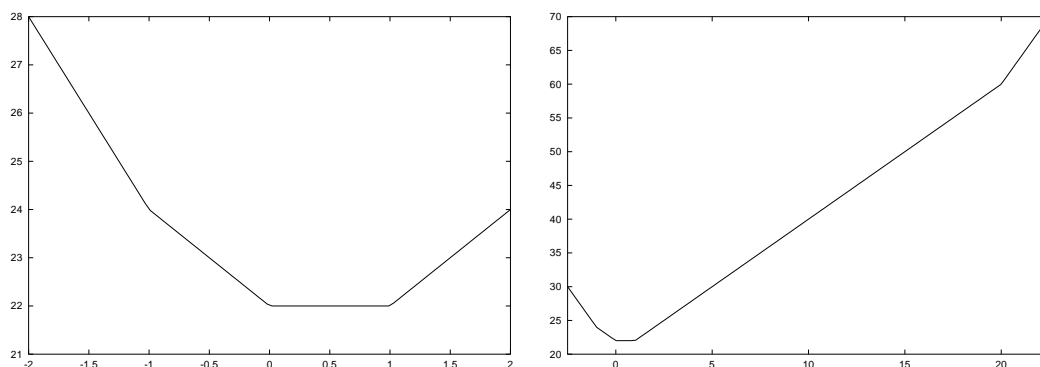


FIG. B.4 – Estimateur L_1 : $res_{\sigma=1}(x)$ sur les intervalles $[-2, 2]$ et $[-2.5, 22.5]$

B.2.2.4 Estimateur L_p

B.2.2.4.1 Présentation Il s'agit ici de présenter les caractéristiques des autres estimateurs L_p , définis dans le tableau B.3 et la figure B.5. Cette famille d'estimateurs n'est intéressante que

$\rho(x)$	ψ	w
$ x ^p/p$	$sgn(x) x ^{p-1}$	$ x ^{p-2}$

TAB. B.3 – Définition de l'estimateur L_p

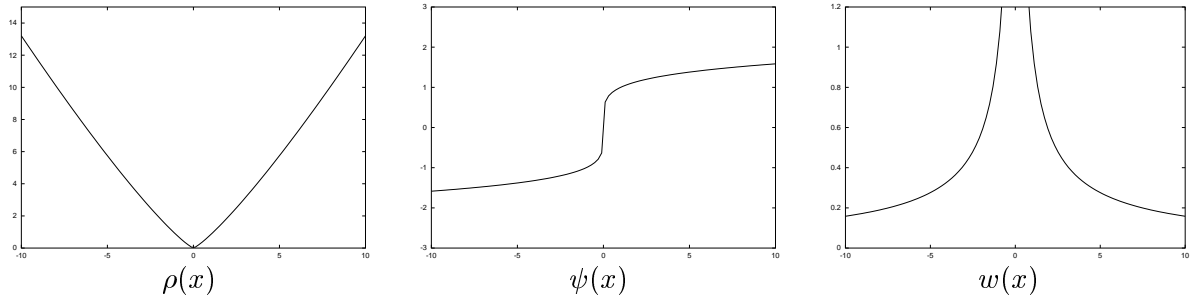
pour $p \in]1, 2[$. En effet :

- $p = 1$ ou $p = 2$: ce sont les estimateurs présentés dans les deux paragraphes précédents ;
- $p > 2$: l'estimateur est plus sensible aux fautes que L_2 , donc non robuste, sans avoir d'autres propriétés intéressantes ;
- $p < 1$: la fonction ρ n'est pas convexe et le comportement proche de la solution devient très problématique.

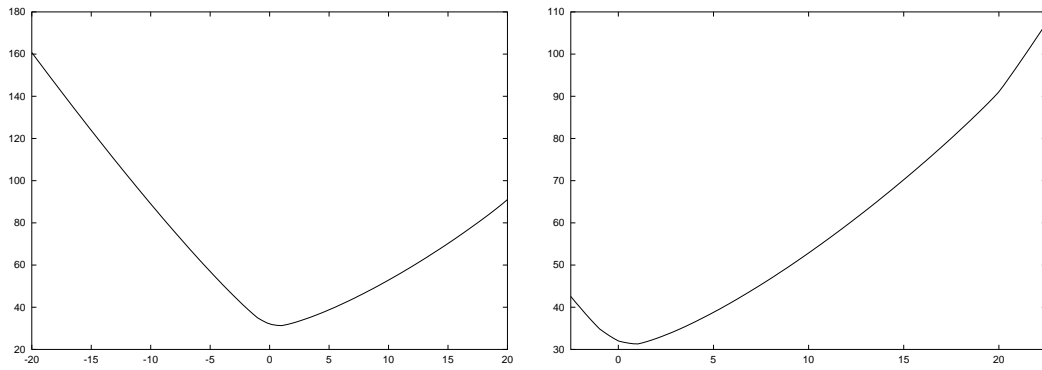
Le choix de la valeur de p résulte d'un compromis :

- plus p est proche de 1, moins les fautes ont d'influence,
- plus p est proche de 2, plus la forme aux environs de la solution s'approche de la forme idéale pour la prise en compte d'un bruit gaussien sur les mesures.

Cependant, pour tout $p \neq 1$, la fonction d'influence ψ n'est pas bornée, et les estimateurs ne sont donc que quasi-robustes. Comme pour l'estimateur L_1 , des problèmes peuvent apparaître lors de la résolution itérative, car les points de résidus nuls induisent des poids infinis, ce qui pose problème. Cependant, de manière pratique, comme la puissance sur le résidu est inférieure à celle de L_1 , la stabilité numérique est bien meilleure. Des études pour la sélection d'un p "optimal" ont été effectuées (cf. [Rey83]), et ont abouties à la valeur $p = 1.2$. Les graphiques des figures B.5 et B.6 ont été réalisés avec la valeur conseillée de $p = 1.2$.

FIG. B.5 – Estimateur L_p : visualisation

B.2.2.4.2 Application La valeur de p retenue est toujours 1.2. Le minimum est atteint pour $x \approx 0.995$ dans tous les cas. Le principal intérêt de cet estimateur est de garantir l'unicité du minimum. On constate que la faute induit un certain décalage de valeur de la solution par rapport à la solution idéale. L'estimation est cependant correcte.

FIG. B.6 – Estimateur L_p : $res_{\sigma=1}(x)$ sur les intervalles $[-2, 2]$ et $[-2.5, 22.5]$

B.2.3 M-estimateurs généraux

B.2.3.1 Généralités

Pour pallier aux défauts des L-estimateurs, un très grand nombre d'estimateurs ont été proposés dans la littérature. Le but est généralement d'avoir un comportement de type L_2 aux abords de la solution, ce qui permet d'avoir le meilleur comportement face au bruit, tout en réduisant, voire en annulant l'influence des fautes. On peut séparer ces estimateurs en deux classes selon leur comportement face aux forts résidus, c'est à dire selon la limite de $\psi(x)$ lorsque $x \rightarrow \infty$:

- ψ croissante et bornée (estimateurs $L_1 - L_2$, Fair et Huber) : ces estimateurs ont un comportement de type L_1 lorsque les résidus deviennent très importants, leurs fonctions d'influence sont bornées, mais les fautes ont tout de même une influence. L'avantage est que la fonction ρ est strictement convexe, ce qui garantit l'unicité du minimum.
- $\psi \rightarrow 0$ (estimateurs Cauchy, Tukey et Welsch) : l'influence des données très éloignées de la solution est nulle ou quasi-nulle. Or, comme ψ est non nulle, elle est forcément décroissante à partir d'un certain point : ρ ne peut être convexe. Il peut donc exister plusieurs minima, ce qui est gênant pour la convergence.

Le problème essentiel de ces estimateurs est le compromis à trouver entre les deux types de comportement (L_1 et L_2), qui nécessite de connaître de manière plus ou moins précise l'écart-type sur les données qui sont justes. Les deux systèmes d'équation (B.13) et (B.14) vont permettre de visualiser le comportement des estimateurs pour une sous-estimation et une sur-estimation de cet écart-type.

Le paragraphe B.2.3.2 présente une méthode de calcul d'une évaluation de l'écart-type des données justes à partir d'une solution approchée.

Les avantages et inconvénients de quelques M-estimateurs courants seront détaillés à partir de la section B.2.3.3.

B.2.3.2 Écart-type robuste

Il est possible, à partir d'une estimation de la solution, d'estimer la valeur de l'écart-type a priori sur les données justes. On note cette estimation $\hat{\sigma}$. La méthode la plus simple, qui est très robuste, consiste à utiliser la médiane des résidus. Dans le cas où le système est très surdéfini, on a par la méthode du maximum de vraisemblance :

$$\hat{\sigma} = \frac{\sqrt{\text{med}_i (r_i^2)}}{0.6745}. \quad (\text{B.21})$$

Dans le cas où le système est peu surdéfini, il est nécessaire d'utiliser la formule :

$$\hat{\sigma} = 1.4826 \left(1 + \frac{1}{m-n} \right) \sqrt{\text{med}_i (r_i^2)}. \quad (\text{B.22})$$

Cette estimation nécessite une solution approchée du problème pour être valable. Elle ne peut donc être utilisée au cours de la résolution du système par les M-estimateurs.

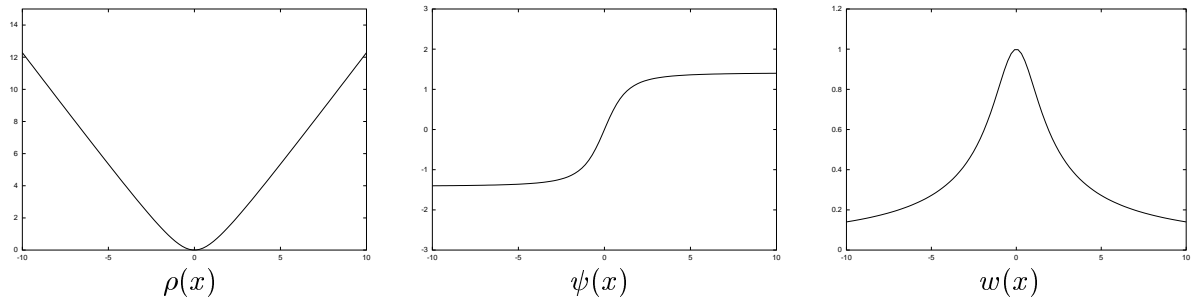
B.2.3.3 Estimateur $L_1 - L_2$

B.2.3.3.1 Présentation Cet estimateur, défini dans le tableau B.2.3.3.1 et la figure B.7, correspond à une fonction simple, C^∞ , ayant un comportement de type L_2 aux alentours de zéro et un comportement de type L_1 lorsque l'on s'en éloigne.

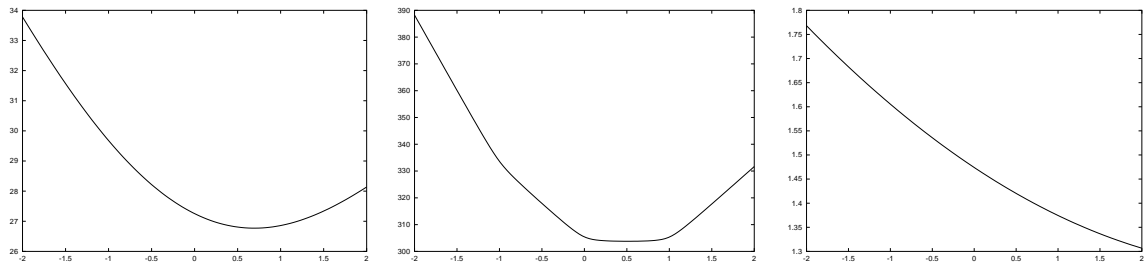
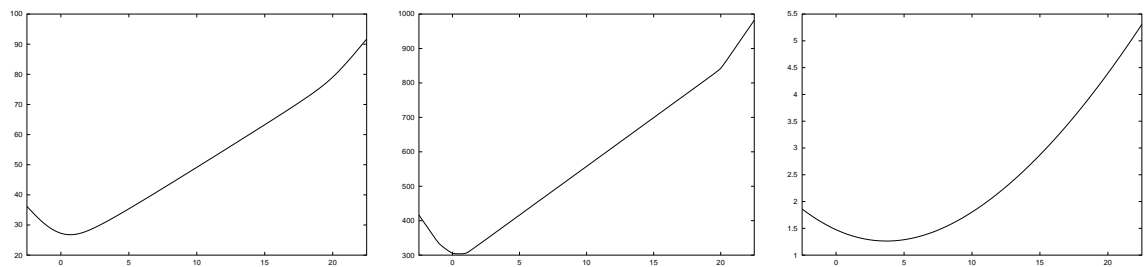
$\rho(x)$	$\psi(x)$	$w(x)$
$2 \left(\sqrt{1+x^2/2} - 1 \right)$	$\frac{x}{\sqrt{1+x^2/2}}$	$\frac{1}{\sqrt{1+x^2/2}}$

TAB. B.4 – Définition de l'estimateur $L_1 - L_2$

En effet, lorsque $x \rightarrow 0$ on a $\sqrt{1+x} \approx 1+x/2$, d'où $\rho(x) \approx 2(1+(x^2/2)/2-1) = x^2/2$, ce qui est bien le comportement de l'estimateur L_2 . De même, lorsque $x \rightarrow \infty$, on a $\rho(x) \approx \sqrt{2}\sqrt{x^2} = \sqrt{2}|x|$, qui a le comportement de l'estimateur L_1 . De plus cette fonction présente l'avantage d'être strictement convexe, ce qui assure l'unicité de la solution. Enfin, la fonction d'influence ψ est bornée : $\forall x, \psi(x) \in]-\sqrt{2}, \sqrt{2}[$.

FIG. B.7 – Estimateur $L_1 - L_2$: visualisation

B.2.3.3.2 Application Le minimum est atteint pour $x \approx 0.695$ dans le premier cas ($\sigma = 1$), $x \approx 0.515$ dans le second ($\sigma = 0.1$) et $x \approx 3.697$ dans le dernier cas ($\sigma = 10$). Lorsque l'écart-type a priori est à peu près exact, cet estimateur fournit une estimation relativement correcte de la solution. Lorsque l'écart-type est sous-évalué, on constate l'apparition d'une cuvette très plate sur l'intervalle $[0, 1]$, qui correspond à un comportement de type L_1 . L'estimateur n'a donc que peu d'intérêt par rapport à l'estimateur L_1 . Inversement, lorsque l'écart-type est très sur-évalué, l'estimateur a un comportement de type L_2 et est donc très influencé par la faute. Il fournit donc un résultat éloigné de la solution idéale.

FIG. B.8 – Estimateur $L_1 - L_2$: $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2, 2]$ FIG. B.9 – Estimateur $L_1 - L_2$: $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2.5, 22.5]$

B.2.3.4 Estimateur Fair

B.2.3.4.1 Présentation Cet estimateur, défini dans le tableau B.5 et la figure B.10, est une fonction C^3 qui garantit l'unicité du minimum.

$\rho(x)$	$\psi(x)$	$w(x)$
$c^2 \left(\frac{ x }{c} - \log \left(1 + \frac{ x }{c} \right) \right)$	$\frac{x}{1+ x /c}$	$\frac{1}{1+ x /c}$

TAB. B.5 – Définition de l'estimateur Fair

Elle utilise les propriétés de la fonction \log pour avoir un comportement de type L_2 aux alentours de 0 et L_1 lorsque les résidus deviennent trop importants. En effet, on a, lorsque $x \rightarrow 0$, $\log(1+x) \approx x - x^2/2$, d'où $|x|/c - \log(1+|x|/c) \approx x^2/(2c^2)$ et donc $\rho(x) \approx x^2/2$. De même, lorsque $x \rightarrow \infty$, on a $x \gg \log(x)$, donc $\rho(x) \approx c|x|$. La constante c , calibrée à partir de la distribution normale (moyenne nulle, $\sigma = 1$), vaut $c = 1.3998$ pour avoir une efficacité de 95%.

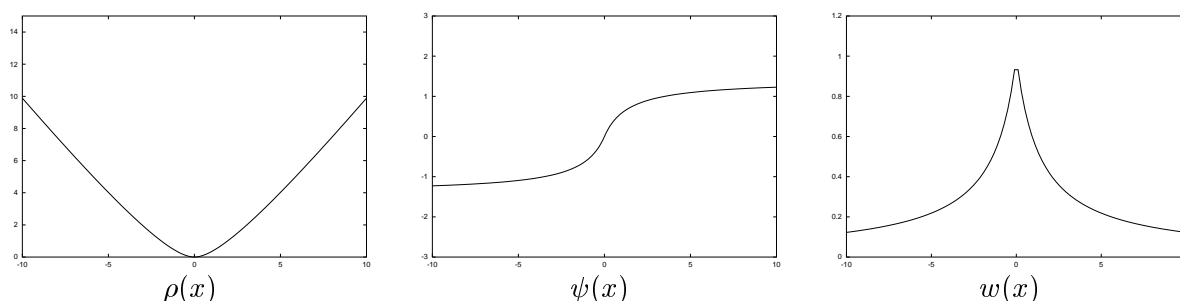
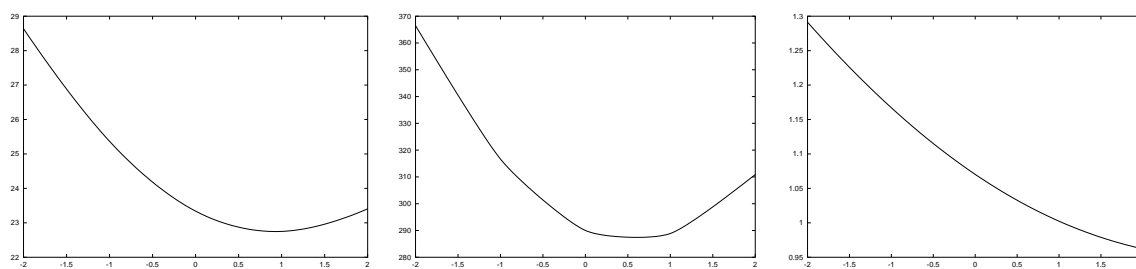
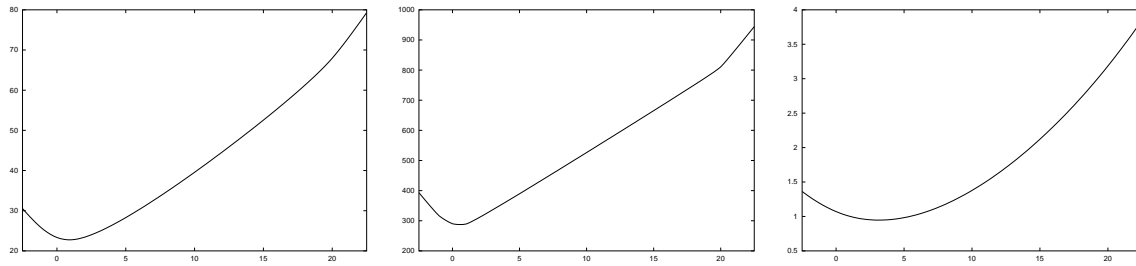


FIG. B.10 – Estimateur Fair : visualisation

B.2.3.4.2 Application Le minimum est atteint pour $x \approx 0.930$ dans le premier cas ($\sigma = 1$). Ainsi, même lorsque l'écart-type a priori est à peu près bien estimé, la faute induit un décalage sensible sur la valeur de la solution. Le minimum est atteint pour $x \approx 0.604$ dans le second cas ($\sigma = 0.1$), et correspond à un comportement de type L_1 sur la zone $[0,1]$. Dans le dernier cas ($\sigma = 10$), le minimum est atteint pour $x \approx 3.154$, donc avec une forte prise en compte de la faute, qui correspond à un comportement de type L_2 . Cet estimateur garantit l'unicité du minimum.

FIG. B.11 – Estimateur Fair : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2,2]$

FIG. B.12 – Estimateur Fair : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2.5, 22.5]$

B.2.3.5 Estimateur Huber

B.2.3.5.1 Présentation L'estimateur de Huber, défini dans le tableau B.6 et la figure B.13, est une fonction définie par morceaux : L_2 en dessous d'une certaine constante et de type L_1 au dessus.

	$\rho(x)$	$\psi(x)$	$w(x)$
si $ x \leq k$	$x^2/2$	x	1
si $ x \geq k$	$k(x - k/2)$	$k \operatorname{sgn}(x)$	$k/ x $

TAB. B.6 – Définition de l'estimateur Huber

La fonction est C^1 , mais malheureusement pas C^2 , ce qui peut entraîner des problèmes de convergence. La constante k , calibrée à partir de la distribution normale (gaussienne de moyenne nulle, $\sigma = 1$), vaut $k = 1.345$ pour avoir une efficacité de 95%.

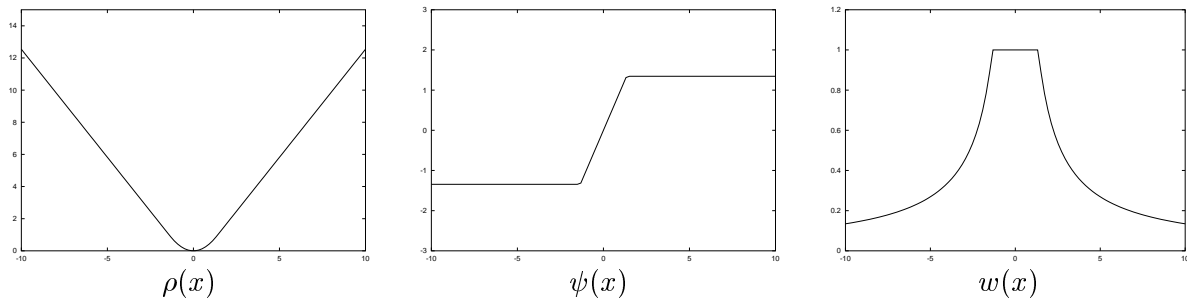
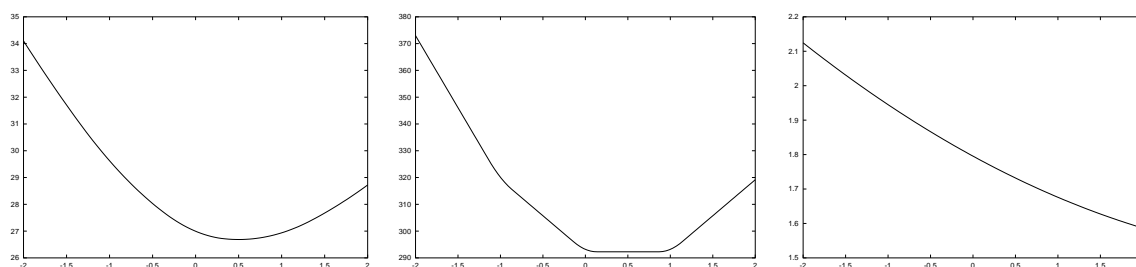
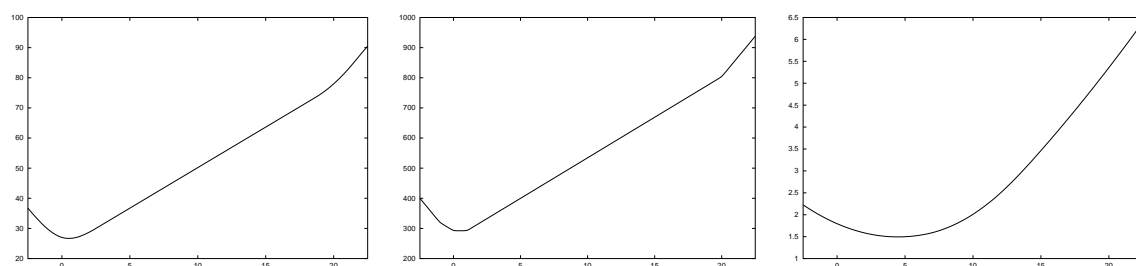


FIG. B.13 – Estimateur Huber : visualisation

B.2.3.5.2 Application Le minimum est atteint pour $x = 0.5$ dans le premier cas ($\sigma = 1$). On constate que la faute induit un faible décalage de valeur de la solution. Dans le second cas ($\sigma = 0.1$), il n'y a pas de minimum strict : sur toute la zone de $x \approx 0.15$ à $x \approx 0.85$ l'estimateur est constant et égal au minimum. En effet, sur cette zone, la somme des écarts est toujours supérieure à la constante k , le comportement est donc en permanence celui de l'estimateur L_1 , à une constante près. Dans le dernier cas ($\sigma = 10$), la solution est $x \approx 4.483$, donc très influencée par la faute. En effet, on est dans le cas inverse du précédent : sur toute la zone $] -2, 22[$ la somme des résidus est inférieure à k , et le comportement est presque celui de l'estimateur L_2 , donc très sensible aux fautes.

FIG. B.14 – Estimateur Huber : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2, 2]$ FIG. B.15 – Estimateur Huber : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2.5, 22.5]$

B.2.3.6 Estimateur Cauchy

B.2.3.6.1 Présentation Cet estimateur, défini par le tableau B.7 et la figure B.16, est également appelé fonction Lorentzienne. Il diminue l'influence des très forts résidus à pratiquement

$\rho(x)$	$\psi(x)$	$w(x)$
$\frac{c^2}{2} \log(1 + (x/c)^2)$	$\frac{x}{1+(x/c)^2}$	$\frac{1}{1+(x/c)^2}$

TAB. B.7 – Définition de l'estimateur Cauchy

0. Il présente le gros inconvénient de ne pas garantir l'unicité du minimum. En effet, il existe des minima locaux, et il est donc possible de tendre vers une solution erronée. La constante c est calibrée en fonction du bruit sur les mesures. Pour une efficacité de 95% sur la distribution normale centrée (gaussienne de moyenne nulle, $\sigma = 1$), la constante vaut $c = 2.3849$.

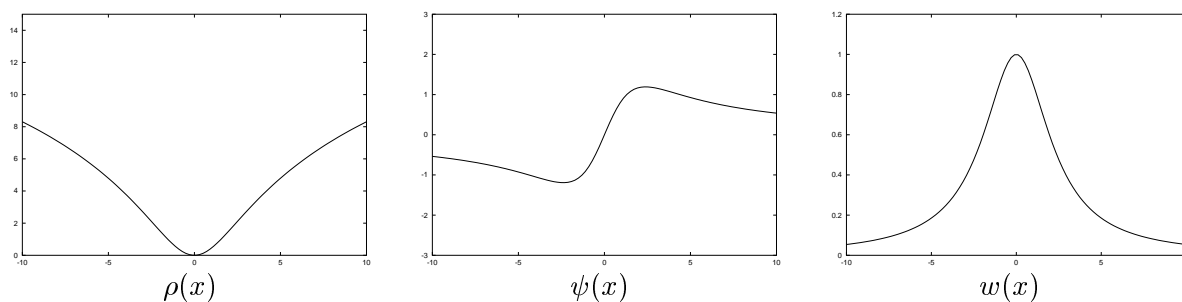


FIG. B.16 – Estimateur Cauchy : visualisation

B.2.3.6.2 Application Dans le premier cas ($\sigma = 1$), le minimum est atteint pour $x \approx 0.129$, le résultat est donc de très bonne qualité. Cependant, sur la vue globale (figure B.18), il apparaît clairement deux minima : cet estimateur peut converger vers une solution complètement fautive. Le second minimum est aux alentours de 20, c'est à dire aux alentours de la donnée fautive ! Dans les cas où l'évaluation de l'écart-type est erronée, c'est encore pire. Dans le second cas ($\sigma = 0.1$), la solution est $x \approx 0.003$, mais il y a 4 minima, correspondant aux 4 données. La convergence devient quasi-impossible à assurer. Dans le dernier cas ($\sigma = 10$), la solution est $x \approx 3.801$, donc très influencée par la faute, car le comportement est pratiquement celui de l'estimateur L_2 .

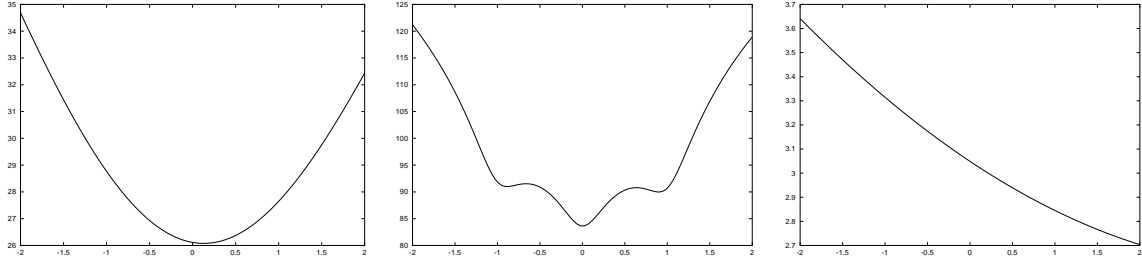


FIG. B.17 – Estimateur Cauchy : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2,2]$

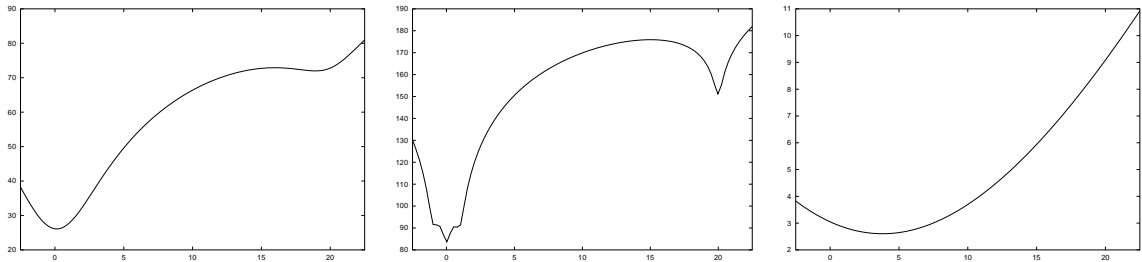


FIG. B.18 – Estimateur Cauchy : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2.5,22.5]$

B.2.3.7 Estimateur Tukey

B.2.3.7.1 Présentation L'estimateur de Tukey, défini par le tableau B.8 et la figure B.19, est une fonction C^3 définie par morceaux.

	$\rho(x)$	$\psi(x)$	$w(x)$
$\begin{cases} x \leq c \\ x > c \end{cases}$	$\begin{cases} c^2/6 \left(1 - [1 - (x/c)^2]^3\right) \\ c^2/6 \end{cases}$	$\begin{cases} x [1 - (x/c)^2]^2 \\ 0 \end{cases}$	$\begin{cases} [1 - (x/c)^2]^2 \\ 0 \end{cases}$

TAB. B.8 – Définition de l'estimateur Tukey

Cette fonction est un polynôme équivalent $x^2/2$ aux alentours de 0, et une constante au-delà d'une certaine valeur. La fonction ψ est donc à support borné : les fautes n'ont pas d'influence. Les problèmes de cet estimateur sont l'existence de multiples minima et l'impossibilité de converger dès que l'on est hors de l'une des cuvettes, puisque $\psi = 0$. La constante c calibrée pour une efficacité de 95% sur la distribution normale vaut $c = 4.6851$.

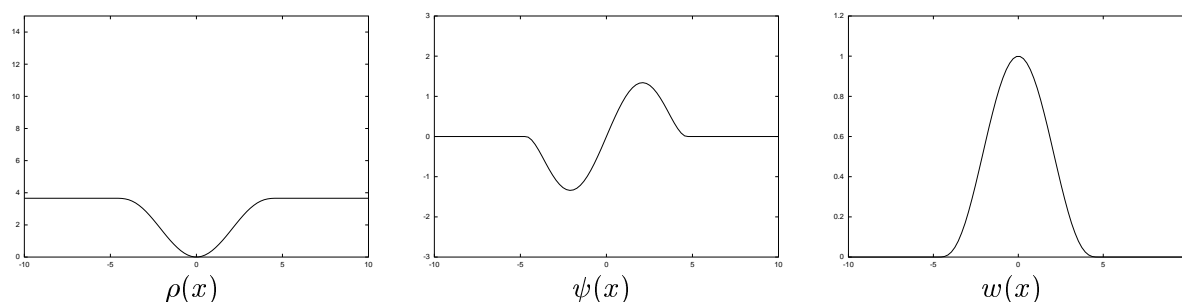
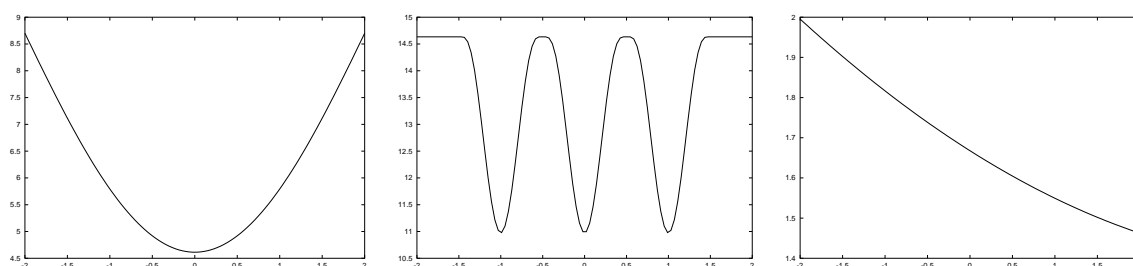
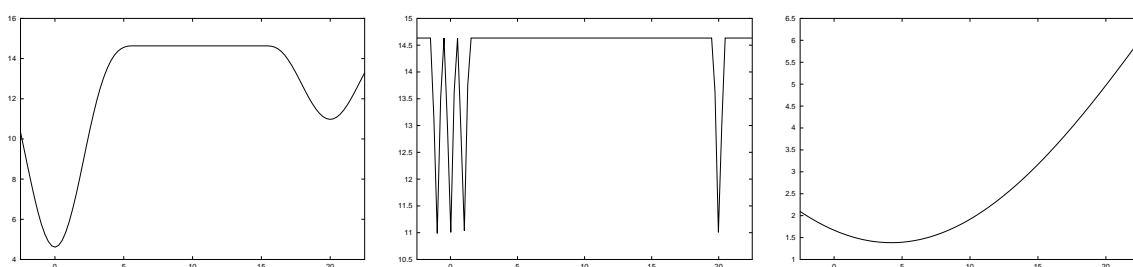


FIG. B.19 – Estimateur Tukey : visualisation

B.2.3.7.2 Application Dans le premier cas ($\sigma = 1$), le minimum est atteint pour $x = 0$, ce qui semble parfait. Cependant, sur la vue globale (figure B.21), il apparaît clairement deux minima : cet estimateur a les mêmes problèmes que l'estimateur de Cauchy : existence de minima locaux et sensibilité à l'écart-type du sur les données sans fautes. Ces deux faiblesses induisent de gros problèmes de convergence. Dans le second cas ($\sigma = 0.1$), on a 4 minima correspondant aux 4 données (-1, 0, 1 et 20), et il n'y a donc même pas de minimum strict. Dans le dernier cas ($\sigma = 10$), la solution est $x \approx 4.211$, c'est à dire un comportement de type L_2 , donc très influencé par la faute.

FIG. B.20 – Estimateur Tukey : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2, 2]$ FIG. B.21 – Estimateur Tukey : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2.5, 22.5]$

B.2.3.8 Estimateur Welsch

B.2.3.8.1 Présentation L'estimateur de Welsch, défini dans le tableau B.9 et la figure B.22, utilise les propriétés de l'exponentielle pour être C^∞ , garantir un comportement de type L_2 aux alentours de 0, et annuler quasiment totalement l'influence des fautes.

$\rho(x)$	$\psi(x)$	$w(x)$
$\frac{c^2}{2} (1 - \exp(-(x/c)^2))$	$x \exp(-(x/c)^2)$	$\exp(-(x/c)^2)$

TAB. B.9 – Définition de l'estimateur Welsch

Le problème principal est l'existence de multiples minima. La constante c , calibrée pour une efficacité de 95% sur la distribution normale vaut $c = 2.9846$.

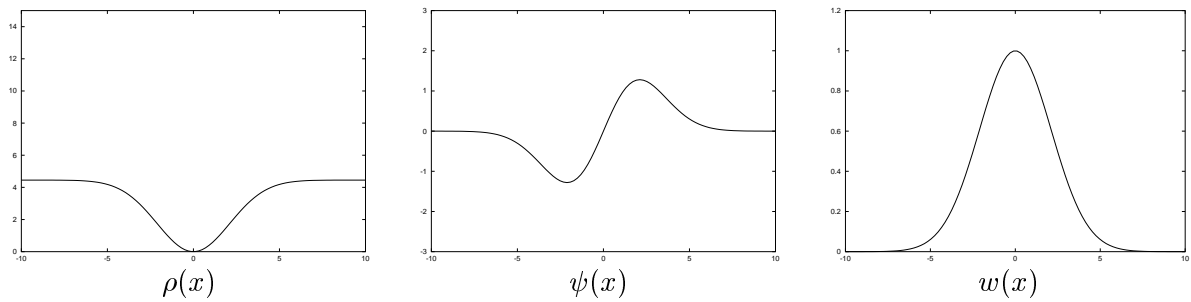
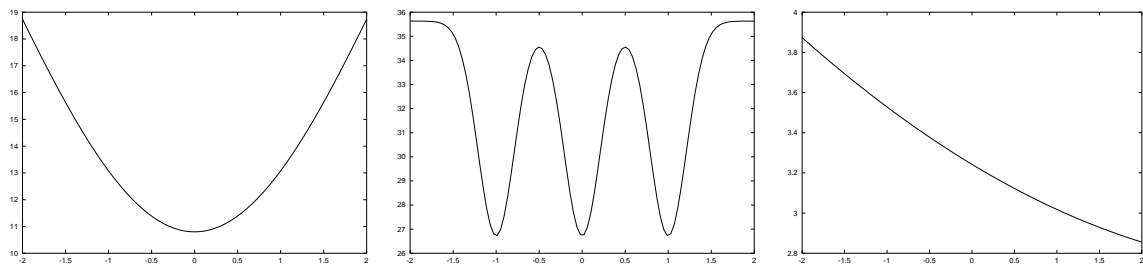
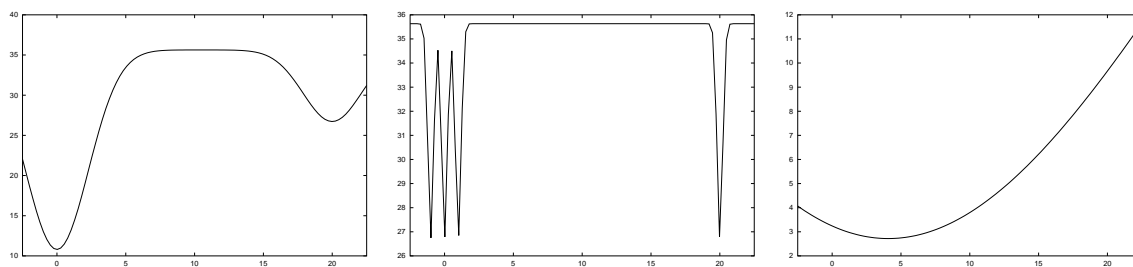


FIG. B.22 – Estimateur Welsch : visualisation

B.2.3.8.2 Application Dans le premier cas ($\sigma = 1$), le minimum est atteint pour $x = 0$, ce qui semble parfait. Cependant, sur la vue globale (figure B.24), il apparaît clairement deux minima. Cet estimateur a les mêmes problèmes que les estimateurs Cauchy et Tukey : même lorsque l'écart-type a priori est bien estimé, il existe des minima locaux. Dans les autres cas, quand l'écart-type est mal estimé, c'est encore plus problématique : dans le cas où $\sigma = 0.1$, il y a 4 minima correspondant aux 4 données et dans le dernier cas ($\sigma = 10$), le minimum est atteint pour $x \approx 4.077$, donc avec une forte prise en compte de la faute.

FIG. B.23 – Estimateur Welsch : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2, 2]$

FIG. B.24 – Estimateur Welsch : $res_{\sigma=1}$, $res_{\sigma=0.1}$ et $res_{\sigma=10}$ sur l'intervalle $[-2.5, 22.5]$

B.2.4 Conclusion

Le choix a priori d'un M-estimateur dans un contexte général est impossible. En effet, en fonction de la connaissance ou non d'un écart-type a priori, d'une solution initiale, l'un ou l'autre des estimateurs présenté aura les propriétés requises.

Lorsqu'il n'y a aucune connaissance a priori sur le problème, il peut être judicieux de combiner les différents estimateurs, afin de profiter de leurs différentes caractéristiques. Ainsi, commencer par l'estimateur L_1 (défini paragraphe B.2.2.3) permet de converger approximativement, sans nécessiter de connaissance sur l'écart-type a priori. Puis, après stabilisation, il est possible de calculer un écart-type sur les données justes de manière robuste (voir paragraphe B.2.3.2). Il suffit alors de terminer par l'un des estimateurs qui s'affranchissent quasi-complètement des fautes (estimateur Welsch par exemple, défini paragraphe B.2.3.8) pour avoir une solution à la fois robuste et précise.

Cependant, cette approche est relativement coûteuse en temps de calcul car la résolution est effectuée plusieurs fois. Or, dans les configurations courantes (en tout cas pour celles que nous avons étudiées), les erreurs ne sont pas complètement aléatoires. Les données peuvent s'éloigner d'un facteur 10 ou 10^2 de leur valeur attendue, mais jamais d'un facteur 10^9 par exemple. Par contre, la connaissance, soit de l'écart-type sur les données justes, soit d'une solution approchée, ne sont pas possibles dans le cas général. L'estimateur L_p a donc été choisi pour nos applications. En effet, il ne nécessite pas de connaissance sur l'écart-type sur les données justes. De plus, compte tenu du type de fautes rencontrées, sa quasi-robustesse ne pose pas de problème. Enfin, par rapport à l'estimateur L_1 , il pose moins de problèmes numériques lors de la résolution itérative (équations à poids infini) et il a un meilleur comportement en présence de bruit gaussien.

C'est donc cet estimateur qui a été retenu pour nos applications, avec la valeur $p = 1.2$.

B.3 Échantillonnage aléatoire

Cette catégorie d'estimateurs robustes s'oppose aux techniques conventionnelles de filtrage de données qui nécessitent une bonne solution initiale pour assurer une convergence vers le modèle correct. Plutôt que d'utiliser toutes les données en même temps, les techniques d'échantillonnage aléatoire essaient de retrouver le modèle correct à partir de petits échantillons de données tirés aléatoirement. Le principe de base des méthodes aléatoires est de trouver, à partir d'un nombre suffisant d'hypothèses, la meilleure d'entre elles, ou celle qui possède la plus grande probabilité de contenir des données correctes. L'espace des solutions est alors restreint à celui des données. Ces techniques sont particulièrement adaptées pour l'estimation de modèles à partir d'un ensemble de données initial de taille importante.

La présentation de ces méthodes s'inspire principalement de la thèse de C. Vestri [Ves00].

B.3.1 LMS : Least Median of Squares

La méthode de la moindre médiane des carrés est représentée comme la résolution du problème de minimisation non linéaire suivant, pour un ensemble de données de n points :

$$\tilde{\mathbf{F}} = \arg \min_F \text{med}_i r_i^2 \text{ pour } i = 0, 1, \dots, n \quad (\text{B.23})$$

La référence principale sur les techniques fondées sur la moindre médiane est le livre de P. Rousseeuw et A. LeRoy [RL87]. La minimisation par la moindre médiane des carrés doit être résolue par la recherche de la solution dans l'espace des données. Le principe consiste à construire des modèles à partir de sous-ensembles de données, puis à évaluer leur qualité de représentativité pour sélectionner le meilleur. On utilise alors un processus de recherche aléatoire qui extrait plusieurs jeux de données servant chacun à construire une hypothèse de modèle. La sélection du modèle solution est réalisée par la recherche de mode de la probabilité de distribution des résidus.

La résolution du système se fait de la manière suivante :

Algorithme 3: Procédure de LMS : Least Median Squares

- 1 **tant que** (*Nombre de tirages* < *seuil*) **faire**
 - Tirer aléatoirement un jeu de données de p points, p étant le nombre minimal de points pour définir un modèle ;
 - Estimer les paramètres de \mathbf{F}_j du modèle ;
 - Calculer les résidus r_i par rapport à ce modèle \mathbf{F}_j , des $n - p$ points restants de l'ensemble des données ;
 - Conserver la valeur médiane des résidus M_j ;
 - Conserver le minimum des valeurs M_j pour obtenir la solution $\tilde{\mathbf{F}}$ qui correspond au modèle associé à cette valeur.
-

Les estimateurs basés sur la moindre médiane des carrés sont biaisés. On utilise souvent un estimateur biaisé pour estimer les valeurs finales des paramètres du modèle. Une estimée robuste de l'écart-type $\hat{\sigma}$ peut être obtenue à partir de la valeur M_j minimale avec les équations B.21 et B.21. L'estimation des paramètres du modèle est alors éalisée par la minimisation des moindres carrés pondérés que l'on notera **RLS : Re-weighted Least Square** :

$$\tilde{\mathbf{F}} = \arg \min_F \sum_i w_i r_i^2 \text{ pour } i = 0, 1, \dots, n \quad (\text{B.24})$$

où les pondérations sont définies par :

$$w_i = \begin{cases} 1 & \frac{r_i}{\hat{\sigma}} \leq 2.5 \\ 0 & \frac{r_i}{\hat{\sigma}} > 2.5 \end{cases} \quad (\text{B.25})$$

B.3.2 Le RANSAC : RANdom SAMple Consensus

Le RANSAC est une procédure exploratoire qui a été introduite par M. Fisher et R. Bolles, dans le traitement d'images en 1981 [FB81]. C'est une technique très utilisée qui s'applique

directement à la reconstruction robuste de scènes. Plusieurs travaux, notamment [SM91, Tor97, Ves00].

Le principe de l'algorithme est le suivant : extraire des modèles à partir des données, déterminer les points qui vérifient chaque modèle et enfin sélectionner la meilleure surface représentant au mieux les données. L'avantage de cette approche est qu'elle permet l'intégration de plusieurs modèles de surface. Pour cela, chaque modèle doit disposer d'une méthode permettant d'évaluer ses paramètres à partir d'un nombre minimal de points p et quantifier la distance d'une donnée par rapport aux modèles calculés. La recherche des points appartenant au modèle est effectuée par croissance de régions. L'algorithme du RANSAC est décrit ci-dessous :

Algorithme 4: Procédure du RANSAC

```

1  tant que (Nombre de modèles < seuil) faire
    - Tirer aléatoirement un type de modèle  $M$  ;
    - Tirer aléatoirement un jeu de données  $S$  de l'ensemble des  $n$  points minimal pour
      déterminer les paramètres du modèle ;
    - Calculer les paramètres du modèle initial  $M^s$  fixé par le jeu  $S$  ;
    - Rechercher dans l'espace  $\{m_i\} - S$  les points qui appartiennent au domaine de
      tolérance du modèle. On obtient un nouveau support  $S^*$  du modèle ;
    si (Nbre de point de  $S^* \geq t$ ) alors
      | le modèle est validé, go to 1 ;
    sinon
      | le modèle est rejeté, go to 1 ;
  
```

Le test d'arrêt est réalisé soit par une estimation du nombre d'itérations nécessaires pour pouvoir retrouver la structure globale de la scène, soit en testant chaque modèle et chaque donnée assez souvent.

Comme nous avons adopté comme unique modèle de surface les facettes, nous n'exploiterons pas la capacité de l'algorithme à intégrer plusieurs modèles. Nous utiliserons donc un modèle unique, le plan. Nous recherchons dans notre cas des hypothèses de plans englobant un nombre suffisant de points.

Les paramètres de l'algorithme sont donc :

- Le domaine de tolérance utilisé pour déterminer si un point est compatible avec le modèle de plan calculé.
- Le critère d'arrêt.
- Le seuil t utilisé pour déterminer si un modèle est représentatif.

Annexe C

Segmentation hiérarchique

C.1 Hiérarchies

C.1.1 Définitions générales

Deux parties A et B sont dites **hiérarchisées** si et seulement si elles sont emboîtées, soit disjointes, c'est-à-dire ssi $A \cap B \in \{A, B, \emptyset\}$ (on dit aussi que A et B n'ont pas d'intersection propre). Par extension, un ensemble H de parties est hiérarchisé si ses éléments le sont tous deux-à-deux.

Si X est un ensemble, une **hiérarchie** sur X est alors un ensemble *hiérarchisé* de parties de X qui vérifie :

- i) $\emptyset \notin H$
- ii) $X \in H$
- iii) $\forall x \in X : \{x\} \in H$.

X est le **sommet** de la hiérarchie. L'ensemble $\mathcal{B}(H) \triangleq \{\{x\}\}_{x \in X}$ est appelée la **base** de la hiérarchie. De façon classique, on assimile une hiérarchie H à une arborescence de racine X notée (H, \succ) , où \succ est la relation de couverture associée à l'inclusion entre parties.

$$x \prec y \Leftrightarrow \{x \subset y \text{ et } \nexists z / x \prec z \prec y\} \quad (\text{C.1})$$

On parlera des noeuds et des arêtes d'une hiérarchie ; les noeuds de base seront appelés les feuilles et le sommet X la racine.

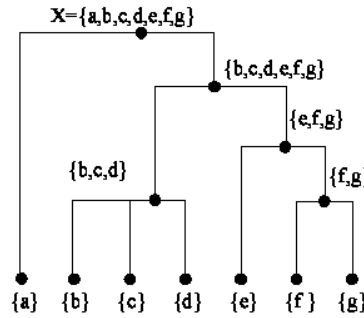


FIG. C.1 – Représentation d'une hiérarchie sous forme de dendrogramme

C.1.2 Père et fils, Branches

Tout élément x de la hiérarchie, à l'exception du sommet, a un unique élément couvrant appelé père de x et noté $p(x)$. Inversement, tout élément x d'une hiérarchie H couvre un certain nombre d'éléments de H , appelés les **fils** de x . Ils sont notés $\mathcal{F}(x) \triangleq \{y \in H, y \prec x\}$.

$\forall x \in H$, la chaîne $x \prec p(x) \prec p(p(x)) \dots \prec X$ est appelée la **branche** de x . Une branche issue d'une feuille de H , qui traverse toute la hiérarchie jusqu'au sommet, est appelée une branche de H .

Le **niveau** d'un élément x dans une hiérarchie et noté $n(x)$ est défini comme la longueur du plus long chemin qui le connecte à la base de la hiérarchie.

C.1.3 Coupe dans une hiérarchie

Du point de vue graphique, une **coupe** de H est un ensemble de noeuds de H que toute branche de H intersecte une seule fois.

Bibliographie

- [Abb94] Y. Abbas. *Base de données vectorielles et erreur cartographique. Problèmes posés par le contrôle ponctuel. Une méthode alternative fondée sur la distance de Hausdorff : le contrôle linéaire*. Thèse de doctorat, Université Paris 7, 1994.
- [AD04] M. Agrawal and L. Davis. Window-based, discontinuity preserving stereo. In *IEEE Conference on Computer Vision and Pattern Recognition CVPR*, pages 66–73, Washington DC, 2004.
- [AF87] N. Ayache and B. Faverjon. Efficient registration of stereo images by matching graph descriptions of edge segment. *Int. Journal of Computer Vision*, 1(2) :107–131, 1987.
- [AF99] B. Ameri and D. Fritsch. 3-d reconstruction of polyhedral-like building models. *IAPRS*, 32(Part 3-2-W5) :15–20, 1999.
- [AFW00] B. Acharya, J. Fagerman, and C. Wright. Accuracy assessment of dtm data : a cost effective approach for a large scale digital mapping project. *International Archives of Photogrammetry and Remote Sensing, XXXIII ISPRS*, 33, Part B2 :105–111, 2000. Amsterdam, Netherlands.
- [Alc01] M. Alcharroun. Positionnement de plans 3d par corrélation multi-vues. In *stage IGN*, 2001.
- [Aya89] N. Ayache. *Vision stéréoscopique et Perception Multisensorielles : Application à la Robotique Mobile*. InterEditions, Paris, 1989.
- [Bad92] A-J. Baddeley. An error metric for binary images. *Robust Computer Vision*, pages 59–78, 1992.
- [Bai97] C. Baillard. *Analyse d'images aériennes stéréoscopiques pour la restitution 3D des milieux urbains*. Thèse de doctorat, ENST, 1997.
- [Bel96] Peter-N. Belhumeur. A Bayesian Approach to Binocular Stereopsis. *International Journal on Computer Vision*, 19(3) :237–262, 1996.
- [BH99] Y. Boykov and D. Huttenlocher. A New Bayesian Framework for Object Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 517–523, 1999.
- [BI99] A-F. Bobick and S-S. Intille. Large Occlusion Stereo. *International Journal of Computer Vision*, 33(3) :181–200, 1999.
- [BJ01] Y. Boykov and M-P. Jolly. Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D images. In *IEEE Conference on Computer Vision*, volume 1, pages 105–112, July 2001.
- [BK04] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/ max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9) :1124–1137, September 2004.

- [BR72] I. Barrodale and F. D. K. Roberts. Solution for an overdetermined system of equations in the L_1 norm. Technical report, University of Victoria, Canada, 1972.
- [BSZF99] C. Baillard, C. Schmid, A. Zisserman, and A. Fitzgibbon. Automatic line matching and 3d reconstruction of buildings from multiple views. In *ISPRS conference on automatic extraction of GIS objects from digital imagery, IAPRS*, volume 32(3), pages 69–80, 1999. Rome, Italie.
- [BT99] S. Birchfield and C. Tomasi. Multiway Cut for Stereo and Motion with Slanted Surfaces. In *IEEE Conference on Computer Vision, ICCV*, September 1999. Greece.
- [BVZ98] Y. Boykov, O. Veksler, and R. Zabih. Markov random fields with efficient approximations. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 648–655, 1998.
- [BVZ99] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. In *ICCV*, volume 1, pages 377–384, 1999.
- [BZ99] C. Baillard and A. Zisserman. Automatic reconstruction of piecewise planar models from multiple views. *CVPR*, pages 559–565, June 1999.
- [CG87] J.P. Cocquerez and A. Gagalowicz. Mise en correspondance de régions dans une paire d'images stéréo. *Actes MARI'87*, pages 166–173, 1987.
- [CJC⁺98] R. Collins, C. Jaynes, Y-Q. Cheng, X. Wang, F. Stolle, E. Riseman, and A. Hanson. The ascender system : Automated site modeling from multiple aerial images. *Computer Vision and Image Understanding*, 72(2) :143–162, November 1998.
- [CJC01] M. Cord, M. Jordan, and J.P. Coquerez. Accurate building structure recovery from high resolution aerial imagery. In *CVIU*, volume 82(2), pages 138–173, May 2001.
- [CJCP99] M. Cord, M. Jordan, J-P. Cocquerez, and N. Paparoditis. Automatic extraction and modeling of urban buildings from high resolution aerial images. *ISPRS*, September 1999. Munich, Germany.
- [Cor98] M. Cord. *Analyse d'images aériennes haute résolution : détection et modélisation du bâti en zone urbaine*. PhD thesis, Université de Cergy-Pontoise, ENSEA-ETIS, 1998.
- [CP97] J-P. Cocquerez and S. Philipp. *Analyse d'images, filtrage et segmentation*. Masson, dunod edition, 1997.
- [Der87] R. Deriche. Using canny's criteria to derive a recursively implemented optimal edge detector. In *International Journal of Computer Vision*, volume 1(2), pages 167–187, May 1987.
- [DKB⁺03] A. Dupéret, M. Kasser, J-Y. Bacon, M. Bernard, and A. Podaire. Observation de la terre : les débuts difficiles mais prometteurs de la résolution sub-métrique optique. *SFPT, colloque Pixels et Cités " Télédétection et photogrammétrie pour le développement en milieu urbain "*, Marne la vallée, 24-26 Novembre, 173/174 (2004-1/2) :3–15, 2003.
- [FB81] M-A. Fischler and R-C. Bolles. Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. In *Communications of ACM*, volume 24, pages 381–385, 1981.
- [FBA01] C-S. Fraser, E. Baltsavias, and A.Grün. 3-d building reconstruction from high resolution ikonos stereo imagery. In E.P. Baltsavias, A. Gruen, and L. Van Gool, editors, *Automated Extraction of Man-Made Objects from Aerial and Space Images (III)*, pages 331–344, Centro Stefano Franscini, Monte Verità, Ascona, June 2001.

- [FC00] J. Fournier and M. Cord. /'evaluation et optimisation d'une chaîne de stéréoresti-
tution par corrélation adaptative. In *RFIA*, Paris, 2000.
- [FF62] L. Ford and D. Fulkerson. *Flows in Networks*. Princeton University Press, 1962.
- [FK98] O. Faugeras and R. Keriven. Complete Dense Stereovision using Level Set Methods.
In *European Conference on Computer Vision (ECCV)*, 1998.
- [FKL⁺98] A. Fischer, T. Kolbe, F. Lang, A. Cremers, W. Förstner, L. Plümer, and V. Stein-
hage. Extracting buildings from aerial images using hierarchical aggregation in 2-d
and 3-d. In *CVIU*, volume 72 :2, pages 163–185, November 1998.
- [FL99] F. Fuchs and H. LeMen. Building reconstruction on aerial images through multi-
primitive graph matching. In *2nd IAPR Work-shop on Graph-based representations*,
Vienna,Austria, 1999. Elsevier Science Ltd.
- [Fua93] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves
image features. In *Machine and Vision Applications*, volume 6, pages 35–49, 1993.
- [Fua97] P. Fua. From Multiple Stereo Views to Multiple 3d-Surfaces. *The International
Journal of Computer Vision*, 24(1) :19–35, 1997.
- [FY97] A. Fu and H. Yan. A new probabilistic relaxation method based on probability
space partition. In *Pattern Recognition*, volume 30 :11. Elsevier Science Ltd, 1997.
- [GBB98] P. Gros, O. Bournez, and E. Boyer. Using local planar geometric invariant to match
and model image of line segments. In *CVIU*, volume 69(2), pages 135–155, February
1998.
- [GC00] M. Gooch and J. Chandler. A novel error detection technique for automatically
generated digital elevation models. *International Archives of Photogrammetry and
Remote Sensing, XXXIII ISPRS*, 33 :302–309, 2000. Amsterdam, Netherlands.
- [GF99] J. Gomes and O. Faugeras. Reconciling distance functions and level sets. Technical
Report 3666, Inria, 1999.
- [GLMC03] L. Guigues, H. Le-Men, and J-P. Cocquerez. Scale-sets image analysis. In *ICIP'03*,
pages 299–306, October 2003. Barcelone,Espagne.
- [GLY95] D. Geiger, B. Ladendorf, and A-L. Yuille. Occlusions and Binocular Stereo. *Inter-
national Journal on Computer Vision*, 14 :211–226, 1995.
- [GLZ00] J. Gong, Z. Li, and Y. Zhou. Effects of various factors on the accuracy of dems :
an intensive experimental investigation. *Photogrammetric Engineering and Remote
Sensing*, 66, No 9 :1113–1117, September 2000.
- [Gui03] L. Guigues. *Modèles multi-échelles pour la segmentation d'images*. Thèse de docto-
rat, Université de Cergy-Pontoise, December 2003.
- [Han89] M-J. Hannah. A system for stereo image matching. *Photogrammetric Engineering
and Remote Sensing*, 55(12) :1765–1770, 1989.
- [HF92] O. Hellwich and W. Faig. Graph-based matching of stereo image features. In *In
XVII Congress of ISPRS, Int. Archives of Photogrammetry and Remote Sensing*,
volume 29, 1992. Rome,Italie.
- [HK86] M. Herman and T. Kanade. Incremental reconstruction of 3d scenes from multiple,
complex images. *Artificial Intelligence*, 30 :289–341, 1986.
- [HS88] C. Harris and M. Stephens. A combined corner and edge detector. In *4th Alvey
Vision Conference*, pages 147–151, 1988.

- [HS89] R. Horaud and T. Skordas. Stereo correspondence through feature grouping and maximal cliques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11) :1168–1180, 1989.
- [IB95] S-S. Intille and A-B. Bobick. Incorporating intensity edges in the recovery of occlusion regions. Technical Report 246, MIT Media Lab Perceptual Computing Group, 1995.
- [IG98a] H. Ishikawa and D. Geiger. Occlusions, discontinuities, and epipolar lines in stereo. In *European Conference on Computer Vision*, pages 232–248, 1998.
- [IG98b] H. Ishikawa and D. Geiger. Segmentation by grouping junctions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 125–131, 1998.
- [Jib02] H. Jibrini. *Reconstruction automatique de bâtiments en modèles polyédriques 3-D à partir de données cadastrales vectorisées 2-D et un couple d'images aériennes à haute résolution*. PhD thesis, ENST, laboratoire MATIS, 2002. IGN-SR-02-005-C-THE-HJ.
- [JPDPM00] H. Jibrini, M. Pierrot-Deseilligny, N. Paparoditis, and H. Maître. Reconstruction de modèles de bâtiments à partir de données cadastrales vectorisées et d'images aériennes. *RFIA*, III :299–306, 2000. Paris, France.
- [KE01a] M. Kasser and Y. Egels. *Photogrammétrie numérique*. Lavoisier, hermes edition, 2001.
- [KE01b] Michel. Kasser and Yves. Egels. *Photogrammétrie numérique (Collection ENSG-IGN)*. Lavoisier, October 2001.
- [KK99] A. Kleusberg and H-G. Klaedtke. Accuracy assessment of a digital height model derived from airborne synthetic aperture radar measurements. In *Photogrammetric week*, 1999.
- [KSC01] S-B. Kang, R. Szeliski, and J. Chai. Handling Occlusions in Dense Multi-View Stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 103–110, 2001.
- [KZ01] Vladimir Kolmogorov and Ramin Zabih. Computing visual correspondence with occlusions via graph cuts. In *ICCV*, pages 508–515, 2001.
- [KZ02] Vladimir Kolmogorov and Ramin Zabih. What energy functions can be minimized via graph cuts? In *ECCV (3)*, pages 65–81, May 2002. Denmark.
- [Lot96] J-L. Lotti. *Mise en correspondance stéréo par fenêtres adaptatives en imagerie haute résolution*. PhD thesis, Université de Nice, Sophia-antipolis, February 1996.
- [LT02] M-H. Lin and C. Tomasi. Surfaces with Occlusions from Layered Stereo. In *CVPR*, 2002.
- [MN89] R. Mohan and R. Nevatia. Using perceptual organisation to extract 3d structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(1) :1121–1139, 1989.
- [MS94] J-C. McGlone and J-A. Shufelt. Projective and Objective Space Geometry for Monocular Building Extraction. In *13th Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pages 54–61, June 1994. Seattle, USA.
- [MT94] S-B. Marapane and M-M. Trivedi. Multi-primitive hierarchical stereo analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3) :227–240, 1994.

- [NH97] C. Brenner N. Haala. Interpretation of urban surface models using 2d building information. *Automated Extraction of Man-Made Objects from Aerial and Space Images (II)*, pages 213–222, May ETH, Ascona, 1997.
- [NN01] S. Noronha and R. Nevatia. Detection and modeling of buildings from multiple aerial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(5) :501–518, May 2001.
- [ODZ03] M. Ortner, X. Descombes, and J. Zerubia. Un nouveau modèle pour l'extraction de caricatures de bâtiments sur des modèles numériques d'élévation. In *TAIMA 03*, pages 213–218, Hammamet, Tunisia, October 2003.
- [OK85] Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 7(2) :139–154, 1985.
- [Pap98] N. Paparoditis. *Reconstruction Tridimensionnelle de Paysages Urbains en Imagerie Stéréoscopique Haute Résolution*. PhD thesis, Université de Nice, Sophia-Antipolis, 1998.
- [PCJC98] N. Paparoditis, M. Cord, M. Jordan, and J-P. Cocquerez. Building detection and reconstruction from mid and high resolution aerial imagery. *Computer Vision and Image Understanding*, 72(2) :122–142, November 1998.
- [Plé05] Pléiades. High resolution panchromatic and multispectral earth observation satellite. <http://smc.cnes.fr/PLEIADES/>, visité le 10/02/2005.
- [Pot52] R. Potts. Some generalized order-disorder transformation. In *the Cambridge Philosophical Society*, volume 48, pages 106–109, 1952.
- [PSQ04] Sylvain Paris, François Sillion, and Long Quan. A surface reconstruction method using global graph cut optimization. In *Proceedings of Asian Conference on Computer Vision*, January 2004.
- [PTC85] T. Poggio, V. Torre, and C.Koch. Computational Vision and Regularization Theory. *Nature*, 317(6035) :314–319, 1985.
- [PTJ00] N. Paparoditis, C. Thom, and H. Jibrini. Surface reconstruction of urban area from highly overlapping digital aerial images. *International Archives of Photogrammetry and Remote Sensing, XXXIII ISPRS*, Geo Information for all, July 2000. Amsterdam, Netherlands.
- [RC98] S. Roy and I. Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *ICCV*, pages 492–502, 1998.
- [RD95] L. Robert and R. Deriche. Dense depth map reconstruction using multiscale regularization approach which preserves discontinuities. In *International Workshop on Stereoscopic and Three-Dimensional Imaging, S. Efstratiadis et al.*, pages 32–39, September 1995.
- [Rey83] W. Rey. *Introduction to robust and quasi-robust statistical methods*. Springer, 1983.
- [RF99] L. Ragia and W. Förstner. Automatically assesing the geometric and structural quality of building ground plans. In *Bulletin SFPT*, volume 153, pages 22–31, 1999.
- [RG92] S. Randriamasy and A. Gagalowicz. Region-based top-down segmentation controlled by stereo matching. *IAPR :MVA1992*, 1992. Japan.
- [RL87] P. J. Rousseeuw and A. M. Leroy. *Robuste Regression and Outlier Detection*. John Wiley & Sons, 1987.

- [SB94] B. Serra and M. Bethod. Subpixel contour matching using continuous dynamic programming. In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pages 202–207, 1994.
- [Sch00] J. Schiewe. Combining geometrical and semantical image information for the improvement of digital elevation model. In *EARSeL Symposium*, 2000.
- [SD01] G. Sohn and I-J. Dowman. Extraction of buildings from high-resolution satellite data. In E.P. Baltsavias, A. Gruen, and L. Van Gool, editors, *Automated Extraction of Man-Made Objects from Aerial and Space Images (III)*, pages 345–355, Centro Stefano Franscini, Monte Verità, Ascona, June 2001.
- [SF01] L-D. Cohen S. Vinson and F. Perlant. Extraction of rectangular buildings using dem and orthoimage. In *SCIA, 12th Scandinavian Conference on Image Analysis*, Bergen, Norway, June 2001.
- [SM91] C-V. Stewart and Minpran. A new robust estimator for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17,10 :925–938, October 1991.
- [SSZ02] D. Scharstein, R. Szeliski, and R. Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1) :7–42, 2002.
- [SVC89] P. Sander, L. Vinet, L. Cohen, and A. Gagalowicz. Hierarchical region based stereo matching. *SCIA, In Proc of the 6th Scandivian Conference on Image Analysis*, pages 71–78, June 1989.
- [SVZ00] D. Snow, P. Viola, and R. Zabih. Exact voxel occupancy with graph cuts. In *Computer Vision and Pattern Recognition Conf*, volume 1, pages 345–353, 2000.
- [SZ97] C. Schmid and A. Zisserman. Automatic line matching across views. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 666–671, 1997.
- [Tai04] F. Taillandier. *Reconstruction de bâti en milieu urbain : une approche multi-vues*. Thèse de doctorat, Ecole Polytechnique, October 2004.
- [TBRN00] B. Thirion, B. Bascle, V. Ramesh, and N. Navab. Fusion of color, shading and boundary information for factory pipe segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 349–356, 2000.
- [Ter86] D. Terzopoulos. Regularization of Inverse Visual Problems involving Discontinuities. *IEEE Transactions on PAMI*, 8(4) :413–424, 1986.
- [TGD03] F. Taillandier, L. Guigues, and R. Deriche. A framework for constrained multi-scale range image segmentation. In *ICIP'03*, October 2003. Barcelone, Espagne.
- [Tor97] P-H-S. Torr. An assessment of information criteria for motion model selection. *Computer Vision and Pattern Recognition*, pages 47–52, 1997.
- [VC02a] S. Vinson and L-D. Cohen. Extraction des bâtiments à partir d'images aériennes et de mne : segmentation of complex buildings from aerial images and dem. In *RFIA*, volume I, pages 125–134, Angers, France, January 2002.
- [VC02b] S. Vinson and L-D. Cohen. Multiple rectangle model for buildings segmentation and 3d scene reconstruction. In *ICPR, 16th Interbational Conference on Pattern Recognition*, volume I, pages 125–134, Québec, Canada, August 2002.
- [VD01] C. Vestri and F. Devernay. Using robust methods for automatic extraction of buildings. In *Proceedings of the International Conference on Computer Vision and Pattern Recongnition, IEEE*, 2001. Hawaii.

- [Vei91] I. Veillet. *Triangulation spatiale de blocs d'images Spot*. PhD thesis, Observatoire de Paris, November 1991.
- [Vek99] O. Veksler. *Efficient Graph-based Energy Minimization Methods in Computer Vision*. PhD thesis, Cornell University, 1999.
- [Vek00] O. Veksler. Image Segmentation by Nested Cuts. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 339–344, 2000.
- [Ves00] C. Vestri. *Outils pour la reconstruction automatique de bâtiments à partir d'images aériennes*. Thèse de doctorat, INRIA, September 2000.
- [Vin91] L. Vinet. *Segmentation et mise en correspondance de régions de paires d'images stéréoscopiques*. Thèse de doctorat, Université de Dauphine, Paris 9, 1991.
- [VT95] J-M. Vézien and J-P. Tarel. A generic approach for planar patches stereo reconstruction. In *Proceedings of the Scandivian Conference on Image Analysis*, pages 1061–1070, 1995. Uppsala, Sweden.
- [WA96] W. Willuhn and F. Ade. A rule-based system for house reconstruction from aerial images. In *ICPR*, pages 885–889, August 1996.
- [Wei97] U. Weidner. Digital surface models for building extraction. *Automated Extraction of Man-Made Objects from Aerial and Space Images (II)*, pages 193–202, May ETH, Ascona, 1997.
- [XZ96] G. Xu and Z. Zhang. *Epipolar Geometry in Stereo, Motion and Object Recognition*. Kluwer Academic Publisher, 1996.
- [ZF92] Z. Zhang and O. Faugeras. *3D scene analysis : a stereo based approach*. Springer, 1992.
- [Zha95] Z. Zhang. Parameter estimation techniques : a tutorial with application to conic fitting. Technical Report 2676, INRIA, October 1995.
- [ZK00] C-L. Zitnick and T. Kanade. A Cooperative Algorithm for Stereo Matching and Occlusion Detection. *IEEE Transactions on PAMI*, 22(7) :675–684, 2000.

Publications

Congrès internationaux avec comité de lecture :

[Chehata et al,05] N. Chehata, M. Pierrot-Deseilligny et G. Stamon. "Hybrid Digital Elevation Models computation constrained by 3D-primitives : A global optimization algorithm using graph cuts". IEEE International Conference on Image Processing ICIP'05, pp 117-120, Septembre 2005, Gènes, Italie.

[CJPDS04] N. Chehata, F. Jung, M.Pierrot-Deseilligny et G. Stamon, "Détection et reconstruction de facettes 3D par approche hiérarchique par régions à partir de couples d'images satellite en milieu urbain", SETIT'04, Mars 2004, Tunisie.

[CJPDS03] N. Chehata, F. Jung, M.Pierrot-Deseilligny et G. Stamon "A Region-Based Matching Approach for 3D-Roof Reconstruction from High Resolution satellite Stereo Pairs", In APRS :DICTA'03, Digital Image Computing Techniques and Applications, vol II, pp 889-898, Décembre 2003, Sydney, Australie.

[CPDJS02] N.Chehata, M. Pierrot-Deseilligny, F.Jung et G.Stamon "Extraction of 3D primitives from stereopairs of satellite images for automatic reconstruction of buildings " In IAPR :MVA2002, p636-639, Décembre 2002, Japon.

Revue avec comité de lecture :

[CPDS,04] N. Chehata, M. Pierrot-Deseilligny et G. Stamon. "Génération de MNE hybrides (raster/ vecteur) : optimisation à base de flots de graphes contrainte par des primitives", revue de la SFPT, n°176(2004-4), pp 30-41, 2004.

Colloques avec actes :

[Che05] N. Chehata "Interprétation de scènes urbaines à partir d'images satellitaires THR : reconstruction de facettes 3D et optimisation globale 3D", Bulletin d'Information de l'IGN, BIST N°75, pp. 29-40.

[GTCTD05] Guigues L., Trias-Sanz R., Chehata N., Taillandier F. et Deveau M., 2005, "Segmentation multi-échelles d'images : théorie et applications", Bulletin d'Information de l'IGN, BIST

N°75, pp. 41-58.